Paper:
van Rijn, E., Carter, N., McMurtrie, H., Willner, P. & Blagrove, M. (2017). Sleep does not cause false memories on a story-based test of suggestibility. *Consciousness and Cognition, 52*, 39-46.
http://dx.doi.org/10.1016/j.concog.2017.04.010

van Rijn, E., Carter, N., McMurtrie, H., Willner, P. & Blagrove, M. T. (2017). Sleep does not cause false memories on a story-based test of suggestibility. *Consciousness and Cognition, 52,* 39–46. doi: 10.1016/j.concog.2017.04.010

Sleep does not cause false memories on a story-based test of suggestibility

Elaine van Rijn[a*], Neil Carter[a], Hazel McMurtrie[a,b], Paul Willner[a] and Mark T. Blagrove[a]
[a]*Department of Psychology, Swansea University, Swansea, United Kingdom*
[b]*Department of Psychology, Manchester Metropolitan University, Manchester, United Kingdom*

*Corresponding author. Current address: Centre for Cognitive Neuroscience, Duke-NUS Medical School, 8 College Road, Singapore, 169857. *E-mail addresses*: e.van-rijn.717567@swansea.ac.uk, elaine.van.rijn@duke-nus.edu.sg (E. van Rijn).

Abstract
Sleep contributes to the consolidation of memories. This process may involve extracting the gist of learned material at the expense of details. It has thus been proposed that sleep might lead to false memory formation. Previous research examined the effect of sleep on false memory using the Deese-Roediger-McDermott (DRM) paradigm. Mixed results were found, including increases and decreases in false memory after sleep relative to wake. It has been questioned whether DRM false memories occur by the same processes as real-world false memories. Here, the effect of sleep on false memory was investigated using the Gudjonsson Suggestibility Scale. Veridical memory deteriorated after a 12-h period of wake, but not after a 12-h period including a night's sleep. No difference in false memory was found between conditions. Although the literature supports sleep-dependent memory consolidation, the results here call into question extending this to a gist-based false memory effect

Keywords: sleep, sleep-dependent memory consolidation, false memory, Gudjonsson Suggestibility Scale

1. Introduction

Sleep contributes to the consolidation of episodic memories (e.g., Weber, Wang, Born, & Inostroza, 2014), as well as declarative memories (e.g., Ellenbogen, Payne, & Stickgold, 2006; Gais & Born, 2004). Sleep-dependent memory consolidation has been argued to be a complex, selective process, during which a process of memory triage takes place to determine which memories are consolidated and in what form (Stickgold & Walker, 2013). According to Stickgold and Walker (2013), during or shortly after encoding memories are tagged for consolidation, and only relevant information is processed during sleep. This results in discriminatory incorporation of selected memories, for example emotional memories (Payne & Kensinger, 2010; Stickgold & Walker, 2013) or memories with future relevance (Fischer & Born, 2009; Saletin, Goldstein, & Walker, 2011; van Dongen, Thielen, Takashima, Barth, & Fernandez, 2012; van Rijn, Lucignoli, Izura, & Blagrove, 2017; Wilhelm et al., 2011), over others. In addition, they propose that new knowledge is formed during two types of sleep-dependent memory integration: item-integration, where newly learned memories are added to existing schemas, and multi-item generalization, where new memories are combined to create a new schema. During multi-item generalization the gist of a set of memories can be extracted, which leads to a combined representation of related memories, while individual items are forgotten (Stickgold & Walker, 2013). In line with the general account of the role of sleep in gist extraction, sleep is proposed to strengthen newly acquired memories through reactivation during slow wave sleep (SWS), until they become integrated with pre-existing knowledge networks in the neocortex (Diekelmann & Born, 2010). The replay of memories during SWS has been proposed to be the underlying mechanism for schema formation and gist extraction (Lewis & Durrant, 2011). During SWS, overlapping replay of memories strengthens areas of similarity (Lewis & Durrant, 2011). Additionally, a process of synaptic downscaling takes place during SWS, in which recently potentiated synapses are downscaled. In this process stronger interconnections between overlapping areas survive, whereas weaker associations are lost (Tononi & Cirelli, 2003). During sleep, shared elements of memories are abstracted from a set of related memories, eventually producing a schema (Lewis & Durrant, 2011). Importantly, it has been suggested that this gist extraction can result in false memories being formed (Lewis & Durrant, 2011; Payne et al., 2009; Stickgold & Walker, 2013).

False memories are recollections of events or stimuli that never actually took place, or which are remembered in a distorted way (Roediger & McDermott, 1995). A commonly used method to research false memories is the Deese-Roediger-McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995). In this paradigm, participants learn lists of semantically related words (e.g., rest, bed, dream, wake) from which a critical theme word, or lure (here, sleep), is omitted. At retest, participants have been found to falsely remember learning the critical lure. Testing is done using either a free recall or a recognition test. In free recall, participants are instructed to recall as many words as they can remember, within a set time frame. In recognition, participants are presented with a series of words, one of which is the lure, and are tested on their recognition of each word (Roediger & McDermott, 1995). Several studies have employed the DRM paradigm to examine whether sleep can cause false memories. In these studies, participants learn multiple DRM lists and are retested after either a period of wake, sleep or sleep deprivation. Results are mixed, with some studies reporting an increase in false memories after sleep (Darsaud et al., 2010; Diekelmann, Born, & Wagner, 2010; McKeon, PaceSchott, & Spencer, 2012; Payne et al., 2009) or sleep deprivation (Diekelmann, Landolt, Lahl, Born, & Wagner, 2008; Diekelmann et al., 2010), whereas others have reported a decrease in false memories after sleep (Fenn, Gallo, Margoliash, Roediger, & Nusbaum, 2009; Lo, Sim, & Chee, 2014), or no effect of sleep on false memory (Cox, Carter, Willner, Blagrove, & van Rijn, 2016).

Sleep is thought to facilitate memory consolidation by extracting the gist of the learned material, in this case the DRM lists, at the expense of the details. This gives rise to a number of mechanisms by which sleep might enhance the recall of false memories. For example, it might be beneficial in some situations to remember the general meaning of information and not all specific details; this would enhance the retrieval of unstudied critical lures because they represent the gist of the DRM list (Lo et al., 2014; Payne et al., 2009). Also, the emotional context may enhance gist memory at the expense of specific details (McKeon et al., 2012). Additionally, the absence of external memory cues during free recall could lead to self-cueing, in which own cues are generated during retrieval so as to remember studied items, which increases the likelihood of falsely remembered unstudied lures as caused by the extraction of generalized features during sleep (Diekelmann et al., 2010). Finally, with free recall, participants need to actively retrieve the studied words, relying more on integrated memory of the lists, which could lead to more spontaneous generation of unstudied items after sleep (McKeon et al., 2012). All of these explanations assume that the processing of memories during sleep results in a gist of the learning being abstracted, with specific and generalized information then being stored, leading to the production of false memories either as part of gist formation, or as a deduction at testing via retrieval of the gist of what was learned (Landmann et al., 2014).

However, it has been questioned whether DRM false memories occur by the same processes as real-world false memories (Frenda, Patihis, Loftus, Lewis, & Fenn, 2014; Ost et al., 2013). Using a richer, more real-world like misinformation task, Frenda et al. (2014) and Lo, Chong, Ganesan, Leong, and Chee (2016) found an increase of false memory after sleep deprivation, but not after sleep. The present study aimed to examine the effect of sleep on false memory in a novel manner that more closely resembles real life, using the story-based non-forensic version of the Gudjonsson Suggestibility Scale (GSS) 2 (Gudjonsson, 1987). The GSS tests for false memory about a story through the use of leading questions, which act as recognition items, but without corresponding veridical information. The GSS was designed to measure susceptibility to suggestion (Gudjonsson, 1984) and has been shown to correlate with real-world instances of false memories or false confessions about criminal acts (Gudjonsson, 1991a,b). The test consists of a short auditory presented story about an incident and a set of 20 questions about this story, 15 of which are leading questions that cannot be answered from the information provided in the story. The other five questions, which are non-leading, are intended to conceal the real purpose of the scale, and are not included in the measurement of suggestibility. Participants listen to the story, recall what they remember from it, and are then tested by answering the 20 questions. After answering all 20 questions, participants receive feedback on their performance. This feedback is typically negative, but can also be neutral (McMurtrie, Baxter, Obonsawin, & Hunter, 2012). After receiving feedback, they answer the questions again. In the present experiment, participants listened to the GSS 2 story and then, 12 h later, after either a period of wake or a period of sleep, answered the questions about it.

Sleep deprived individuals have been found to be more suggestible on the GSS than controls who are not sleep deprived (Blagrove, 1996). This may indicate that sleep deprivation leads to a failure to discriminate between the original story and the suggestive questions. Alternatively, sleep deprived participants could be less motivated and confident in their ability to answer the questions correctly (Blagrove, 1996). This second explanation was tested in a subsequent study, in which participants gave confidence ratings on a scale of 1 (extremely uncertain) to 5 (extremely certain) after answering each question. Sleep deprived participants were again significantly more suggestible, but their levels of confidence were the same as for the controls. It was thus proposed that the effect is not caused by motivational deficits, but by cognitive deficiencies in source memory due to effects of sleep loss on the prefrontal cortex (Blagrove & Akehurst, 2000).

Mazzoni (2002) distinguish between ‘naturally occurring’ and ‘suggestion-dependent’ memory distortions. Naturally occurring memory errors can arise at any time and result from the way memory works. Suggestion-dependent memory distortions, on the other hand, occur due to the presence of external suggestions, such as leading questions. DRM errors are a type of naturally occurring memory errors, whereas GSS errors are suggestion-dependent (Mazzoni, 2002). Given that sleep has been shown in some studies to result in naturally occurring DRM false memories (e.g., Darsaud et al., 2010; Diekelmann et al., 2010; McKeon et al., 2012; Payne et al., 2009), our aim in this study is to test whether suggestion-dependent false memories are augmented by sleep.

Whereas the DRM paradigm has been used to examine the effect of sleep, sleep deprivation and wake on false memory, sleep and non-sleep deprived wake groups have not been compared directly on the GSS before now. The GSS uses a narrative, the gist of is hypothesized to be consolidated during sleep. The use of leading questions about the narrative is proposed to elicit a false memory response. In free recall of DRM list words, where no cues are presented, the unstudied critical lures are falsely recalled, as they represent the gist of the DRM list. During a DRM recognition test, on the other hand, the veridical cues act as reminders of the details, and therefore may not bring up the false memory of the gist word. When comparing the GSS with the DRM paradigm, the use of leading questions in the GSS might initially seem more similar to a DRM recognition task rather than a free recall task. However, the GSS leading questions do not act as cues in the same way as DRM list words do during recognition. The GSS leading questions act as cues by referring to elements from the story, however, this is to enable a plausible false memory to be suggested. These leading questions therefore differ from a DRM recognition test where the veridical cues are actual learned list words. The GSS leading questions are akin to presenting just the lure words in a recognition test on the DRM paradigm. It was hypothesized that gist processes occurring during sleep would lead to increased susceptibility to answering the leading questions incorrectly during wakefulness the following day. This hypothesis was examined both for overall scores on the 15 leading questions and for a subgroup of the leading questions that are more strongly premised in the story. In addition, it was expected that veridical memory of the story content would be better after a period of sleep than after a period of wake, also due to sleep-dependent memory consolidation processes.

2. Methods

2.1. Participants

Ninety-four participants (39 male, 55 female) aged 18–53 (mean = 25.10, SD = 9.38) volunteered to take part in the experiment. All participants were native English speakers. Participants self-reported usually sleeping ≥6 h per night and had no history of sleep disorders, memory disorders, neurological, or psychiatric disorders. Participants were asked to refrain from alcohol intake for one day before and during the experiment. Participants were randomly assigned to the wake group or sleep group. Fifty-one participants comprised the wake group (mean age = 24.22, SD = 8.49), and 43 participants the sleep group (mean age = 26.14, SD = 10.34). Next, participants were randomly assigned to the neutral or negative feedback condition. Forty-five participants were assigned to the neutral feedback condition (mean age = 25.49, SD = 9.95) and 49 participants to the negative feedback condition (mean age = 24.73, SD = 8.91). The number of participants as a function of group, feedback condition and gender can be found in Table 1. All participants gave written informed consent and the experiment was approved by the Research Ethics Committee of the Swansea University Department of Psychology.

Table 1

Number of participants as a function of group, feedback condition and gender

	Wake		Sleep		Total		Overall total
	Male	Female	Male	Female	Male	Female	
Neutral feedback	11	17	6	11	17	28	45
Negative feedback	10	13	12	14	22	27	49
Total	21	30	18	25	39	55	
Overall total	51		43				

2.2. Materials

2.2.1. The Stanford Sleepiness Scale

All participants completed the Stanford Sleepiness Scale (SSS) (Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973) at the start of session 1 and session 2. This scale assesses how alert one is feeling at that moment on a scale of 1 (feeling active, vital, alert, or wide awake) to 7 (no longer fighting sleep, sleep onset soon; having dream-like thoughts). The SSS (Hoddes et al., 1973) is a frequently used subjective measurement of sleepiness (Bailes et al., 2006). It is sensitive in revealing daytime sleepiness in normal subjects, though it might not be suitable to use in pathological populations (Curcio, Casagrande, & Bertini, 2001). The SSS is sensitive in measuring sleepiness following total and partial sleep deprivation (Herscovitch & Broughton, 1981; Hoddes et al., 1973). Furthermore, the SSS has been found to correlate with the Profile of Mood States (POMS), three sleepiness related Visual Analogue Scales (VAS) (Pilcher, Pury, & Muth, 2003) and two fatigue scales (Bailes et al., 2006).

2.2.2. The Gudjonsson Suggestibility Scale

The non-forensic version of the Gudjonsson Suggestibility Scale (GSS) 2 (Gudjonsson, 1987) comprises a short spoken story about a couple named Anna and John, who witness a boy going down a steep slope on a bicycle, calling for help. Participants give immediate free recall and delayed free recall (here, 12 h later) for the story. For each correctly remembered "idea" from the story (parsed according to semantic structure), participants score one point. A total of 40 points can be scored for this free recall. After the delayed free recall five non-leading and 15 leading questions about the story are responded to, and scored for the number of correct (for non-leading questions) or affirmative (for leading questions) responses. The leading and non-leading questions are interspersed.

The GSS distinguishes between three types of leading question: those questions that create expectations by including salient premises from the story, those that have an affirmative response bias, and those that use false alternatives to imply the presence of objects, persons and events that are not mentioned in the story (Gudjonsson, 1984). The first category of leading question here is the most similar to the DRM critical lures, in that DRM lures are plausible expectations from DRM lists. Four leading questions were identified by consensus among the research team as most similar conceptually to the DRM critical lures, as they create expectations by including salient premises from the story (questions 4, 12, 18 and 20, see below for questions and relevant story content). This set of four questions was analyzed separately in addition to the overall analysis of affirmative answers to all 15 leading questions.

4. Story content: 'John worked in a bank.' Question: Was the husband a bank director?

12. Story content; '[…] John caught hold of the bicycle and brought it to a halt.' Question: Did John grab the boy's arm or shoulder?

18. Story content: 'The boy appeared very frightened, but unhurt […].' Question: Was the boy frightened of riding the bicycle again?

20. Story content: 'Sometimes in the winter months the two couples had gone skiing together […].' Question: Did the couple have a skiing cottage in the mountains?

2.2.3. Feedback

After answering all 20 questions, participants received feedback on their performance, which could be negative or neutral. The negative feedback was taken from Gudjonsson (1987) and stated: 'You have made a number of errors. It is therefore necessary to go through the questions once more, and this time try to be more accurate.' The neutral feedback was taken from McMurtrie et al. (2012) and stated: 'Thank you for answering these questions. To ensure that we have your answers recorded correctly, we'll run through the questions once more.' Participants answered the same 20 questions again after receiving one of the two types of feedback. Assignment to feedback condition was randomized prior to participants taking part, and hence was unrelated to scores on the first presentation of GSS questions.

The second presentation of questions was used so as to elicit higher false memory scores, and possibly higher false memory scores for the sleep group in comparison to the wake group, given their hypothesized susceptibility to gist errors. The GSS also distinguishes between shift and total suggestibility. Shift refers to the number of answers that are changed from the first to the second round of questions, as a result of interrogative pressure from the feedback given in-between the two rounds of questions. Total suggestibility is the sum of the number of incorrectly answered questions before feedback and the number of shifted responses. The shift and total suggestibility measures thus have an interpersonal pressure aspect, which is not so relevant to the ideas of memory consolidation and were therefore not included in the analyses.

2.2.4. Memory confidence

After responding to each question participants were asked how confident they are in their answer on a scale from 0 to 100 percent. They received the following instructions: 'Please type a number from 0% to 100% to show how certain you are of what you wrote down, where 0% = completely uncertain and 100% = completely certain.' A textbox was provided at the end of each question response box for participants to enter the percentage.

2.3. Procedure

All participants participated in two sessions, presented online in their own home, and spaced exactly 12 h apart for each participant. Participants in the sleep group completed session 1 of the experiment at 20.00–22.00 in the evening and were tested exactly 12 h later the next morning for session 2, whereas participants in the wake group completed session 1 at 08.00–10.00 in the morning and were then tested exactly 12 h later in the evening for session 2. Before each session, participants were asked how many hours they had slept in the previous 12 h, to ensure that participants in the wake group had not napped in-between the two sessions, and that the participants in the sleep group had slept at least six hours in-between the two sessions. In session 1, participants listened to the GSS short story presented once as an audio stream from the experiment website. After listening to the story, they were instructed to type out as much as they could remember from the story into a text box (immediate free recall). In session 2, participants first typed out in a text box what they could remember from the story they had listened to twelve hours earlier (delayed free recall). Next, participants answered the GSS questions, which comprised the interspersed five non-leading questions and the 15 leading questions, the latter to measure false memory before feedback. After answering the 20 questions participants received feedback, after which they answered the 20 questions again, presented in the same order.

2.4. Statistical analyses

To measure veridical story memory before and after sleep or wake, the scores on the immediate and delayed free recall were calculated for each participant. False memory before and after feedback was calculated for both the sleep and wake group, based on how many of the 15 leading questions were answered affirmatively. In addition, the responses to the four questions that were conceptually most similar to the DRM critical lures were analyzed separately. The five non-leading, factual questions about the story were scored for the number of correct responses. Finally, for both groups, mean confidence ratings (in percentage) were calculated for correctly and incorrectly answered leading questions, and non-leading questions, both before feedback and after feedback.

The effect of sleep on the different memory measures was first analyzed using 2 × 2 Repeated-Measures ANOVAs. The two groups (sleep versus wake) were compared on their immediate and delayed free recall, followed by paired-samples *t*-tests where appropriate. (In one instance, planned comparisons were conducted in the absence of a significant interaction on the basis of the a priori hypothesis.) The two groups were also compared on number of affirmative responses to leading questions, and confidence, before and after feedback, followed by paired-samples t-tests where appropriate. For effect size measures partial eta squared (η_p^2) was calculated for ANOVA results and Cohen's *d* for *t*-test results.

Next, JASP (Love et al., 2015) was used to perform Bayesian Repeated-Measures ANOVAs, Bayesian paired-samples and independent samples t-tests, as well as Bayesian contingency tables, to estimate the Bayes factors of the memory measures, false memory measures and the confidence ratings. For Bayesian Repeated-Measures ANOVAs, the following default priors were used: *r* scale fixed effects = 0.5, *r* scale random effects = 1, *r* scale covariates = 0.354. For Bayesian *t*-tests, a Cauchy prior distribution was used (width = 0.71). The Bayesian approach can be used to calculate the probability, or likelihood, of a hypothesis given the data (Dienes, 2011). The Bayes factor (BF) indicates the ratio of likelihoods of the data supporting the null hypothesis or the alternative hypothesis. When comparing the null hypothesis against the alternative hypothesis (BF_{01}), a greater Bayes factor indicates support for the null hypothesis; when comparing the alternative hypothesis against the null hypothesis (BF_{10}), a greater BF indicates support for the alternative hypothesis (Dienes, 2011). For example, in BF_{01} a Bayes factor of 1 to 3 indicates some evidence for the null hypothesis and a Bayes factor greater than 3 indicates substantial evidence for the null hypothesis, whereas a Bayes factor of less than 1 supports the alternative hypothesis. A Bayes factor of 1 is inconclusive (Dienes, 2011; Wetzels, Grasman, & Wagenmakers, 2012). Given the mixed results of previous research on the effect of sleep on false memory using the DRM paradigm, the Bayesian approach was considered useful to examine whether the null or the alternative hypothesis was supported by the current data.

3. Results

Mean time differences in decimal hours between session 1 and session 2 for the two groups are presented in Table 2. The wake and sleep groups did not differ in time between the two sessions ($t(92) = 0.10$, *ns*). Mean scores on the Stanford Sleepiness Scale for the two groups are also presented in Table 2. There was no significant main effect of session ($F(1, 92) = 2.18$, ns), no significant main effect of group ($F(1, 92) = 0.004$, *ns*), and no significant interaction between session and group ($F(1, 92) = 3.22$, *ns*).

Table 2

Means (SDs) of testing-related variables for the wake and sleep groups

Measure	Wake	Sleep
Mean time between sessions [a]	11.99 (0.28)	11.99 (0.23)
Mean sleepiness score at learning [b]	2.47 (1.03)	2.81 (1.43)
Mean sleepiness score at testing [b]	2.53 (1.23)	2.21 (1.17)

Note. [a] Mean time between start of session 1 (learning) and start of session 2 (testing) in decimal hours, [b] Stanford Sleepiness Scale score.

3.1. Memory and false memory measures

Means and standard deviations of story recall and false memory for both groups are presented in Table 3. A significant main effect of time on story recall (immediate versus delayed free recall) was found ($F(1, 92) = 10.61$, $\eta_p^2 = 0.10$, $p = .002$, $BF_{10} = 20.80$), indicating higher overall scores on immediate free recall compared to delayed free recall. The main effect of group ($F(1, 92) = 0.07$, *ns*, $BF_{01} = 1.73$) and the interaction effect were not significant ($F(1, 92) = 0.04$, *ns*, $BF_{01} = 4.02$), but paired-samples *t*-tests, testing the a priori predicted decrease in veridical memory following the 12 hour period of wake, but not sleep, showed that the wake group remembered significantly less from the story after the 12 hour delay ($t(50) = 2.83$, $p = .007$, $d = 0.21$, $BF_{10} = 5.28$), whereas there was no significant change in free recall for the sleep group ($t(42) = 1.82$, *ns*, $BF_{01} = 1.34$).

Concerning the main hypothesis, there was no significant main effect of group on false memory (F(1, 90) = 0.02, *ns*, BF_{01} = 2.34), and no significant interaction between group and pre-/post feedback factor (F(1, 90) = 0.48, *ns*, BF_{01} = 3.35). The four questions that were conceptually most similar to the DRM critical lures, due to having highest premise expectancies, were analysed separately. No significant main effect of group on answers to these four questions was found (F(1, 90) = 0.24, *ns*, BF_{01} = 1.91), no significant main effect of pre-/post-feedback factor (F(1, 90) = 1.35, *ns*, BF_{01} = 4.15) was found, and there was no significant interaction of group with pre-/post-feedback factor (F(1, 90) = 1.65, *ns*, BF_{01} = 1.19).

For answers to the five non-leading, factual questions about the story there was no significant main effect of group (F(1, 90) = 1.93, *ns*, B_{F01} = 0.73), no significant main effect of feedback (F(1, 90) = 0.35, *ns*, BF_{01} = 6.00), and no significant interaction between group and pre-/post-feedback factor (F(1, 90) = 0.85, *ns*, BF_{01} = 2.76).

There was a significant increase in false memory after feedback (F(1, 90) = 12.98, p < .01, η_p^2 = 0.13, BF_{10} = 11.57), and the interaction between pre-/post-feedback test and feedback type was significant (F(1, 90) = 17.05, p < .001, η_p^2 = 0.16, BF_{10} = 518.29). Paired-samples t-tests indicated that false memory increased significantly after receiving negative feedback (t(48) = -4.51, p < .001, d = 0.34, BF_{10} = 505.80), but not after receiving neutral feedback (t(44) = 0.94, *ns*, BF_{01} = 4.08). Importantly, feedback type did not interact with group (sleep versus wake) on false memory (F(1, 90) = 0.67, *ns*, BF_{01} = 1.60). Feedback type was included in the analyses where appropriate, but the effect of feedback type will not be reported here, as it does not alter the findings or conclusions.

Table 3

Means (SDs) of the GSS memory and false memory measures for the wake and sleep groups

Measure	Wake	Sleep
Immediate free recall score	18.27 (5.86)	17.00 (5.85)
Delayed free recall score	17.06 (5.89)	16.16 (5.73)
False memory responses before feedback	5.88 (3.59)	6.14 (3.62)
False memory responses after feedback	6.59 (4.15)	6.78 (4.01)
False memory responses on DRM-like questions before feedback	1.69 (1.22)	1.98 (1.14)
False memory responses on DRM-like questions after feedback	1.84 (1.33)	1.93 (1.24)
Score on non-leading questions before feedback	4.49 (0.70)	4.33 (0.99)
Score on non-leading questions after feedback	4.57 (0.81)	4.28 (0.98)

3.2. Memory confidence

Means and standard deviations of confidence ratings (in percentage) for the two groups are presented in Table 4. Memory confidence of the sleep and wake groups was analysed separately for answers to leading and non-leading questions before and after feedback. No group difference in confidence was found for answers to leading questions ($F(1, 90) = 0.11$, *ns*, $BF_{01} = 2.02$). For non-leading questions, participants were significantly more confident in their answers after feedback compared to before feedback ($F(1, 90) = 13.48$, $p < .001$, $\eta_p^2 = 0.13$, $BF_{10} = 33.61$), but there was no difference between the sleep and wake groups ($F(1, 90) = 0.26$, *ns*, $BF_{01} = 2.32$). All two-way interactions were not significant (all *p*s > .05, all $BF_{01} > 3$).

No differences in confidence for (correct) non-affirmative responses to leading questions were found between groups ($F(1, 90) = 0.27$, *ns*, $BF_{01} = 3.64$) or for before compared to after feedback ($F(1, 90) = 0.03$, *ns*, $BF_{01} = 3.86$). Participants were significantly more confident in their incorrectly answered leading questions after feedback than before feedback ($F(1, 83) = 4.92$, $p = .03$, $\eta_p^2 = 0.06$, $BF_{10} = 0.22$), but no differences were found between groups in confidence for (incorrect) affirmative responses to leading questions ($F(1, 83) = 0.13$, *ns*, $BF_{01} = 2.70$). All two-way interactions were not significant (all *p*s > .05, all BF_{01} > 4).

No differences in confidence for (correct) non-affirmative answers to the leading questions conceptually most similar to the DRM lures were found between groups ($F(1, 77) = 0.0002$, *ns*, $BF_{01} = 3.48$) and there was no difference between pre-and post-feedback responses ($F(1, 77) = 2.07$, *ns*, $BF_{01} = 2.95$). Participants were significantly more confident in their incorrectly answered DRM-like leading questions after feedback than before feedback ($F(1, 73) = 4.73$, $p = .03$, $\eta_p^2 = 0.06$, $BF_{10} = 0.23$). No group differences were found in confidence for (incorrect) affirmative responses to the leading questions conceptually most similar to the DRM lures ($F(1, 73) = 0.60$, *ns*, $BF_{01} = 2.88$). All two-way interactions were not significant (all *p*s > .05, all BF_{01} > 4).

Table 4

Means (SDs) of confidence percentage in response to GSS questions for the wake and sleep groups as a function of type of question and type of response

Measure	Before feedback		After feedback	
	Wake	Sleep	Wake	Sleep
Mean confidence on leading questions	65.15 (17.57)	63.00 (20.21)	66.21 (22.39)	64.16 (23.63)
Mean confidence on non-leading questions	90.36 (11.35)	88.75 (10.92)	92.50 (12.49)	91.70 (9.25)
Mean confidence on affirmative responses to leading questions	59.69 (15.21)	57.81 (21.20)	64.36 (21.56)	61.41 (22.65)
Mean confidence on non-affirmative responses to leading questions	64.87 (22.11)	61.46 (24.07)	64.74 (24.24)	61.47 (29.19)
Mean confidence on affirmative responses to DRM-like questions	69.69 (24.25)	68.39 (24.98)	69.20 (19.33)	68.47 (20.93)
Mean confidence on non-affirmative responses to DRM-like questions	66.13 (26.45)	63.28 (28.24)	69.25 (17.75)	62.98 (22.46)

4. Discussion

The present study investigated the effect of sleep on the formation of false memories, using the complex materials from the Gudjonsson Suggestibility Scale. A significant decrease in story memory was found across wake, but not across sleep. This finding accords with previous literature and supports a memory consolidation or other protective effect of sleep (e.g., Ellenbogen et al., 2006; Gais & Born, 2004). For both the sleep and wake group, there was a significant increase in false memory after negative feedback, which is consistent with previous research (e.g., McMurtrie et al., 2012). However, the hypothesized gist-based false memory effect due to sleep did not occur in the experiment here. There was also no false memory effect of sleep for the four questions from the GSS that are conceptually most similar to the DRM paradigm. The Bayes factors (BF01) comparing the sleep and wake groups on the GSS false memory measures ranged from 1.19–3.35, indicating “some to substantial” evidence for the null hypothesis that there is no effect of sleep on false memory (Dienes, 2011; Wetzels et al., 2012). Furthermore, no differences were found between the sleep and wake groups in confidence ratings and the Bayes factors (BF01), ranging from 2.70–4, indicated “some to mainly substantial” evidence for the null hypothesis (Dienes, 2011; Wetzels et al., 2012) of no effect of sleep on confidence in responses. These findings indicate the absence even of any sleep-related gist or false-memory processes that might have been evidenced implicitly by confidence rather than by behavior.

The present findings are not in line with previous studies that found an increase of false memories after sleep using the DRM paradigm (Diekelmann et al., 2010; McKeon et al., 2012; Payne et al., 2009). However, not all previous studies found that sleep leads to an increase in false memories. Table 5 summarizes seven previous studies of false memory after sleep using the DRM paradigm. The table is organized according to whether memory was tested using free recall or recognition. Some evidence of the effect of sleep on false memory was seen in all studies that used free recall (Diekelmann et al., 2010; McKeon et al., 2012; Payne et al., 2009), but more mixed results, including both increases (Darsaud et al., 2010) and decreases (Fenn et al., 2009; Lo et al., 2014) in false memory, have been reported in studies using recognition. Lo et al. (2014) found decreased false memory after sleep compared to wake, with longer duration of slow wave sleep (SWS) being correlated with decreased number of false memories after sleep. They propose a mechanism that synaptic connections for contextual details are weaker for false than for veridical memories, resulting in false memory connections being pruned during SWS, whereas contextual and sensory information related to veridical memories have stronger synaptic connections and are less likely to be removed during synaptic downscaling. Their results and the behavioral data here do support the standard memory consolidation view of the strengthening and interconnecting of memories during sleep, but the extrapolation to a gist account, with false memories resulting, is not supported by the present findings or by those of Lo et al. (2014).

It is unlikely that the absence of a false memory effect is caused by circadian effects in this experiment. Participants in the wake and sleep group did not differ on sleepiness during session 1 or session 2. Furthermore, no time of day effects were found by three others studies that did report an effect of sleep on false memories using the DRM paradigm (Fenn et al., 2009; McKeon et al., 2012; Payne et al., 2009).

Various types of consolidation and processing of memories have been proposed to occur during sleep. These range from selective consolidation of emotional experiences, about which there is considerable evidence (Payne & Kensinger, 2010; Stickgold & Walker, 2013), schema formation (Lewis & Durrant, 2011), and selective consolidation according to future relevance (Fischer & Born, 2009; Saletin et al., 2011; van Dongen et al., 2012; van Rijn et al., 2017; Wilhelm et al., 2011), to arguably more sophisticated processes of insight discovery, creativity and gist-based false memories, for which the evidence is mixed or more sparse (Landmann et al., 2014; Stickgold & Walker, 2013). The findings from the experiment here support the view that sleep can ameliorate the decrease in declarative memory across time that occurs during wake, but suggest that the processes of sleep-dependent memory consolidation might not extend to the gist-based formation of narrative or suggestion-dependent false memories (Mazzoni, 2002). The hypothesized effect of sleep on false memory has been discussed in several recent review papers. Some reviews acknowledge that the mixed findings of previous studies make it difficult to characterize the sleep-dependent consolidation processes as potentially involved in the formation of false memories (Rasch & Born, 2013; Stickgold & Walker, 2013; Straube, 2012). However, other reviews do not discuss all research done in the area, but instead cite only the studies showing an increase in false memory after sleep (Born & Wilhelm, 2012; Conte & Ficca, 2013; Stickgold, 2013). The present negative findings, and the lack of clarity in the literature, indicate that more replications are needed (see Open Science Collaboration, 2015) to examine claims that sleep increases false memory, using both the DRM paradigm and more real-world stimuli, such as the story-based GSS. When taken together with the mixed results from the DRM studies, the present results call into question the reliability and generalizability of the sleep and false memory effect.

Table 5

Summary of previous DRM studies on sleep or sleep deprivation and false memory

Authors	Participants	DRM Paradigm	Main findings
		Free recall	
Payne et al. (2009)	14-43 per group (S and W)	8 lists: 12 words per list	Increased FM after S compared to W; increased VM after S compared to W.
Diekelmann et al. (2010)	18-19 per group (S, SD and W)	8 lists: 12 words per list	Increased FM after S and SD compared to W, but only with low general memory performance; no difference in VM.
McKeon et al. (2012)	15-17 per group (S and W)	Emotional DRM with 10 lists: 10 words per list (5 neutral and 5 negative)	Increased FM after S compared to W for both emotional and neutral words; increased VM after S compared to W.
		Recognition	
Diekelmann et al. (2008)	14-18 per group (S and SD)	18 lists: 15 words per list	Increased FM after SD compared to S, effect disappears with caffeine before retrieval; no difference in VM.
Fenn et al. (2009)	46 per group (S and W)	10-16 lists: 15 words per list	Decreased FM after S compared to W; no difference in VM.
Darsaud et al. (2010)	18 per group (S and SD)	32 lists: 15 words per list	Increased FM after S compared to SD; increased VM after S compared to SD.
Lo et al. (2014)	14 (S and W, within subjects)	10 lists: 15 words per list	Decreased FM after S compared to W; no difference in VM.

Note. S=sleep; SD= sleep deprivation; W=wake; FM=false memory; VM=veridical memory

5. References

Bailes, S., Libman, E., Baltzan, M., Amsel, R., Schondorf, R., & Fichten, C. S. (2006). Brief and distinct empirical sleepiness and fatigue scales. *Journal of Psychosomatic Research, 60*(6), 605-613. doi: 10.1016/j.jpsychores.2005.08.015

Blagrove, M. (1996). Effects of length of sleep deprivation on interrogative suggestibility. *Journal of Experimental Psychology: Applied, 2*(1), 48-59. doi: 10.1037/1076-898X.2.1.48

Blagrove, M., & Akehurst, L. (2000). Effects of sleep loss on confidence-accuracy relationships for reasoning and eyewitness memory. *Journal of Experimental Psychology: Applied, 6*(1), 59-73. doi: 10.1037/0278-7393.6.1.59

Born, J., & Wilhelm, I. (2012). System consolidation of memory during sleep. *Psychological Research, 76*(2), 192-203. doi: 10.1007/s00426-011-0335-6

Conte, F., & Ficca, G. (2013). Caveats on psychological models of sleep and memory: A compass in an overgrown scenario. *Sleep Medicine Reviews, 17*(2), 105-121. doi: 10.1016/j.smrv.2012.04.001

Cox, E., Carter, N., Willner, P., Blagrove, M. T., & van Rijn, E. (2016). Sleep does not cause false memories on the Deese-Roediger-McDermott paradigm. *Journal of Sleep Research, 25*(Issue Supplement S1), 205. doi: 10.1111/jsr.12446

Curcio, G., Casagrande, M., & Bertini, M. (2001). Sleepiness: Evaluating and quantifying methods. *International Journal of Psychophysiology, 41*, 251-263. doi: 10.1016/S0167-8760(01)00138-6

Darsaud, A., Dehon, H., Lahl, O., Sterpenich, V., Boly, M., Dang-Vu, T., . . . Collette, F. (2010). Does sleep promote false memories? *Journal of Cognitive Neuroscience, 23*(1), 26-40. doi: 10.1162/jocn.2010.21448.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*(1), 17-22. doi: 10.1037/h0046671

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience, 11*(2), 114-126. doi: 10.1038/nrn2762

Diekelmann, S., Born, J., & Wagner, U. (2010). Sleep enhances false memories depending on general memory performance. *Behavioural Brain Research, 208*(2), 425-429. doi: 10.1016/j.bbr.2009.12.021

Diekelmann, S., Landolt, H. P., Lahl, O., Born, J., & Wagner, U. (2008). Sleep loss produces false memories. *PLOS ONE, 3*(10), 1-9. doi: 10.1371/journal.pone.0003512.g001

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Psychological Science, 6*(3), 274–290. doi: 10.1177/1745691611406920

Ellenbogen, J. M., Payne, J. D., & Stickgold, R. (2006). The role of sleep in declarative memory consolidation: Passive, permissive, active or none? *Current Opinion in Neurobiology, 16*(6), 716-722. doi: 10.1016/j.conb.2006.10.006

Fenn, K. M., Gallo, D. A., Margoliash, D., Roediger, H. L., III, & Nusbaum, H. C. (2009). Reduced false memory after sleep. *Learning & Memory, 16*(9), 509-513. doi: 10.1101/lm.1500808

Fischer, S., & Born, J. (2009). Anticipated reward enhances offline learning during sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(6), 1586-1593. doi: 10.1037/a0017256

Frenda, S. J., Patihis, L., Loftus, E. F., Lewis, H. C., & Fenn, K. M. (2014). Sleep deprivation and false memories. *Psychological Science*, 1-8. doi: 10.1177/0956797614534694

Gais, S., & Born, J. (2004). Declarative memory consolidation: Mechanisms acting during human sleep. *Learning & Memory, 11*(6), 679-685. doi: 10.1101/lm.80504

Gudjonsson, G. H. (1984). A new scale of interrogative suggestibility. *Personality and Individual Differences, 5*(3), 303-314. doi: 10.1016/0191-8869(84)90069-2

Gudjonsson, G. H. (1987). A parallel form of the Gudjonsson Suggestibility Scale. *British Journal of Clinical Psychology, 26*, 215-221. doi: 10.1111/j.2044-8260.1987.tb01348.x

Gudjonsson, G. H. (1991a). The effects of intelligence and memory on group differences in suggestibility and compliance. *Personality and Individual Differences, 12*(5), 503-505. doi: 10.1016/0191-8869(91)90070-R

Gudjonsson, G. H. (1991b). Suggestibility and compliance among alleged false confessors and resisters in criminal trials. *Medicine, Science, and the Law, 31*(2), 147-151. doi: 10.1177/002580249103100210

Herscovitch, J., & Broughton, R. (1981). Sensitivity of the Stanford Sleepiness Scale to the effects of cumulative partial sleep deprivation and recovery oversleeping. *SLEEP, 4*(1), 83-92.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of sleepiness: A new approach. *Psychophysiology, 10*(4), 431-436. doi: 10.1111/j.1469-8986.1973.tb00801

Landmann, N., Kuhn, M., Piosczyk, H., Feige, B., Baglioni, C., Spiegelhalder, K., . . . Nissen, C. (2014). The reorganisation of memory during sleep. *Sleep Medicine Reviews, 18*(6), 531-541. doi: 10.1016/j.smrv.2014.03.005

Lewis, P. A., & Durrant, S. J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends in Cognitive Sciences, 15*(8), 343-351. doi: 10.1016/j.tics.2011.06.004

Lo, J. C., Chong, P. L., Ganesan, S., Leong, R. L., & Chee, M. W. (2016). Sleep deprivation increases formation of false memory. *Journal of Sleep Research*, 1-10. doi: 10.1111/jsr.12436

Lo, J. C., Sim, S. K., & Chee, M. W. (2014). Sleep reduces false memory in healthy older adults. *SLEEP, 37*(4), 665-671. doi: 10.5665/sleep.3564

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., . . . Wagenmakers, E.-J. (2015). JASP (Version 0.7.1.12) [Computer software].

Mazzoni, G. (2002). Naturally occurring and suggestion-dependent memory distortions: The convergence of disparate research traditions. *European Psychologist, 7*(1), 17-30. doi: 10.1027//1016-9040.7.1.17

McKeon, S., Pace-Schott, E. F., & Spencer, R. M. (2012). Interaction of sleep and emotional content on the production of false memories. *PLOS ONE, 7*(11), e49353. doi: 10.1371/journal.pone.0049353

McMurtrie, H., Baxter, J. S., Obonsawin, M. C., & Hunter, S. C. (2012). Consistent witness responses: The effects of age and negative feedback. *Personality and Individual Differences, 53*(8), 958-962. doi: 10.1016/j.paid.2012.07.008

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. doi: 10.1126/science.aac4716

Ost, J., Blank, H., Davies, J., Jones, G., Lambert, K., & Salmon, K. (2013). False memory not equal false memory: DRM errors are unrelated to the misinformation effect. *PLOS ONE, 8*(4), e57939. doi: 10.1371/journal.pone.0057939

Payne, J. D., & Kensinger, E. A. (2010). Sleep's role in the consolidation of emotional episodic memories. *Current Directions in Psychological Science, 19*(5), 290-295. doi: 10.1177/0963721410383978

Payne, J. D., Schacter, D. L., Propper, R. E., Huang, L. W., Wamsley, E. J., Tucker, M. A., . . . Stickgold, R. (2009). The role of sleep in false memory formation. *Neurobiology of Learning and Memory, 92*(3), 327-334. doi: 10.1016/j.nlm.2009.03.007

Pilcher, J. J., Pury, C. L. S., & Muth, E. R. (2003). Assessing subjective daytime sleepiness: An internal state versus behavior approach. *Behavioral Medicine, 29*, 60-67. doi: 10.1080/08964280309596058

Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Reviews, 93*, 681-766. doi: 10.1152/physrev.00032.2012

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. doi: 10.1037/0278-7393.21.4.803

Saletin, J. M., Goldstein, A. N., & Walker, M. P. (2011). The role of sleep in directed forgetting and remembering of human memories. *Cerebral Cortex, 21*(11), 2534-2541. doi: 10.1093/cercor/bhr034

Stickgold, R. (2013). Parsing the role of sleep in memory processing. *Current Opinion in Neurobiology, 23*(5), 847-853. doi: 10.1016/j.conb.2013.04.002

Stickgold, R., & Walker, M. P. (2013). Sleep-dependent memory triage: Evolving generalization through selective processing. *Nature Neuroscience, 16*(2), 139-145. doi: 10.1038/nn.3303

Straube, B. (2012). An overview of the neuro-cognitive processes involved in the encoding, consolidation, and retrieval of true and false memories. *Behavioral and Brain Functions, 8*, 35. doi: 10.1186/1744-9081-8-35

Tononi, G., & Cirelli, C. (2003). Sleep and synaptic homeostasis: A hypothesis. *Brain Research Bulletin, 62*(2), 143-150. doi: 10.1016/j.brainresbull.2003.09.004

van Dongen, E. V., Thielen, J. W., Takashima, A., Barth, M., & Fernandez, G. (2012). Sleep supports selective retention of associative memories based on relevance for future utilization. *PLOS ONE, 7*(8), e43426. doi: 10.1371/journal.pone.0043426 van Rijn, E., Lucignoli, C., Izura, C., & Blagrove, M. T. (2017). Sleep-dependent memory consolidation is related to perceived value of learned material. *Journal of Sleep Research*, 26. doi: 10.1111/jsr.12457

Weber, F. D., Wang, J. Y., Born, J., & Inostroza, M. (2014). Sleep benefits in parallel implicit and explicit measures of episodic memory. *Learning & Memory, 21*(4), 190-198. doi: 10.1101/lm.033530.113

Wetzels, R., Grasman, R. P. P. P., & Wagenmakers, E.-J. (2012). A default bayesian hypothesis test for ANOVA designs. *The American Statistician, 66*(2), 104-111. doi: 10.1080/00031305.2012.695956

Wilhelm, I., Diekelmann, S., Molzow, I., Ayoub, A., Molle, M., & Born, J. (2011). Sleep selectively enhances memory expected to be of future relevance. *Journal of Neuroscience, 31*(5), 1563-1569. doi: 10.1523/JNEUROSCI.3575-10.2011