Note

Sleep preserves original and distorted memory traces

Scott A. Cairney a, Shane Lindsay b, Ken A. Paller c and M. Gareth Gaskell a,∗

a Department of Psychology, University of York, York, UK
b Psychology, School of Life Sciences, University of Hull, Hull, UK
c Department of Psychology, Northwestern University, Evanston, IL, USA

Abstract

Retrieval facilitates the long-term retention of memories, but may also enable stored representations to be updated with new information that is available at the time of retrieval. However, if information integrated during retrieval is erroneous, future recall can be impaired: a phenomenon known as retrieval-induced distortion (RID). Whether RID causes an “overwriting” of existing memory traces or leads to the co-existence of original and distorted memory traces is unknown. Because sleep enhances memory consolidation, the effects of sleep after RID can provide novel insights into the structure of updated memories. As such, we investigated the effects of sleep on memory consolidation following RID. Participants encoded word locations and were then tested before (T1) and after (T2) an interval of sleep or wakefulness. At T2, the majority of words were placed closer to the locations retrieved at T1 than to the studied locations, consistent with RID. After sleep compared with after wake, the T2-retrieved locations were closer to both the studied locations and the T1-retrieved locations. These findings suggest that RID leads to the formation of an additional memory trace that corresponds to a distorted variant of the same encoding event, which is strengthened alongside the original trace during sleep. More broadly, these data provide evidence for the importance of sleep in the preservation and adaptive updating of memories.

© 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Retrieval practice can greatly benefit long-term retention, but it may also enable stored representations to be updated with new information that is available at the time of retrieval (Dudai & Eisenberg, 2004). Memory updating via retrieval provides a means of adapting to changes in the external environment. However, if retrieved information is erroneous, its integration within an existing representation can cause distortion and impair future recall (Schacter, Guerin, & St Jacques, 2011). Multiple bouts of remembering in the
absence of suitable feedback and correction may therefore lead to substantial memory inaccuracies.

Such retrieval-induced distortion (RID) was demonstrated by Bridge and Paller (2012). Participants learned a set of object locations and were then tested on all objects immediately after training, on a subset of objects the following day, and again for all objects another day later. RID was observed as locations recalled in the final test were preferentially closer to the locations recalled the prior day than to the locations that were originally learned. Moreover, brain potentials at the time of intervening retrieval predicted later memory distortion. Thus, rather than simply strengthening memory for the studied information, retrieval promoted the storage of retrieved information, which was often inaccurate.

Standard models of systems consolidation propose that episodic memories are initially dependent on both the hippocampus and neocortex, but become gradually independent of the hippocampus as consolidation unfolds (Marr, 1971; McClelland, 2013; McClelland, McNaughton, & O’Reilly, 1995). The multiple trace-transformation (MTT) account of memory, by contrast, proposes that the hippocampus is involved in the retrieval of episodic memories for as long as they exist (Moscovitch & Nadel, 1998; Nadel & Moscovitch, 1997; Winocur, Moscovitch, & Bontempi, 2010; Winocur, Moscovitch, & Sekeres, 2013). A central tenet of this model is that each time an episodic memory is retrieved it is re-encoded within the hippocampus as a new trace with unique contextual properties. Thus, the more often a memory is retrieved, the greater the number of corresponding hippocampal traces that will exist. Neocortical networks may therefore draw upon these traces to support a gist-like, decontextualised version of the original memory. From an MTT perspective, RID may lead to the presence of an additional hippocampal memory trace that corresponds to a distorted variant of the same encoding event. Subsequent retrieval operations would then draw upon both the original and distorted traces, resulting in a blend of the two. Alternatively, if the hippocampus does not re-encode distinct episodic traces during retrieval, then memory updating and distortion may result from some form of “overwriting” within relevant neural networks. Retrieval operations in this framework would then utilise only a single, distorted trace during memory recall.

There is now robust evidence that sleep facilitates the consolidation of hippocampal-dependent, episodic spatial memories (Rasch, Buchel, Gais, & Born, 2007; Rudy, Voss, Westerberg, & Paller, 2009; Wilhelm, Diekelmann, & Born, 2008). How sleep influences originally-learned and distorted spatial memories can thus provide novel insights into the neurocognitive mechanisms of memory updating. Accordingly, we examined the effects of sleep on memory consolidation following RID. Participants encoded word locations and were then tested before (T1) and after (T2) a period of sleep or wakefulness. RID was indicated when T2-recalled locations were closer to T1-recalled locations than to the studied locations. Thus, the distance between locations recalled at T1 and T2 provided an index of distorted memory content. However, the change in word-location error (distance between the studied and recalled locations) from T1 to T2 provided an index of original memory content.

We predicted three possible outcomes. First, if memory updating is achieved by an “overwriting” of existing information at retrieval, then sleep after RID should strengthen distorted but not original memory content (i.e., better recall of T1 locations but not studied locations after sleep vs wake). Second, if RID leads to an additional memory trace for the distorted location, then sleep should strengthen this and the original location trace to similar extents (i.e., better recall of both T1 and studied locations after sleep vs wake). Third, if post-retrieval memory traces are in some way subsidiary to original memory traces, then sleep should preferentially strengthen the original location memories (i.e., better recall of the studied locations than the T1 locations after sleep vs wake).

2. Materials and methods

2.1. Participants

Sixty healthy males were randomly assigned to a sleep group (n = 30, mean ± SD age = 19.77 ± 1.33 years) or a wake group (n = 30, mean ± SD age = 20.10 ± 1.49 years). Participants had no history of sleep, psychiatric or neurological disorders, were medication-free and had not consumed alcohol/caffeine within 24 h of the study.

2.2. Procedure

Two sessions were separated by a 120-min interval (see Fig. 1A). The first session commenced at 9pm (enabling the sleep group to have a normal bed time), and began with a word-location task. For each 3 sec trial of an initial passive viewing phase, participants viewed a rectangular box containing one of 50 words from Maki, McKinley, and Thomson (2004). Words were presented on a grid background and appeared in randomised screen locations. After two passive viewing rounds, participants carried out an active learning phase. For each trial, one of the 50 words was presented centrally and, using the mouse, participants moved it to the location studied at passive viewing. The word then reappeared in the studied (i.e., correct) location for 3 sec. After two active learning rounds, words that had been consecutively placed within a standardised distance of 150 pixels (4.8 cm on our 27" 1920 × 1080 display) from the studied location were dropped from the task and active learning continued until this criterion was met for all remaining words (Bridge & Paller, 2012; Cairney, Lindsay, Sobczak, Paller, & Gaskell, 2016). There was a marginal group difference in the number of rounds required to reach the criterion for all 50 words [sleep group mean ± SD = 10.00 ± 2.95 rounds; wake group mean ± SD = 12.00 ± 4.50 rounds; t(58) = 2.04, p = .05]. Importantly, however, there were no significant group differences in memory performance at baseline (see Section 3). The first of two tests (T1) for all 50 word locations was then carried out. This followed the same procedures as one round of active learning, but without location feedback. As a secondary assessment of declarative memory, participants then completed a word pair task. Details of this task and the results are available in the Supplementary Materials. Afterwards (~11pm), sleep group participants went to bed for 90 min. Sleep was confirmed with polysomnography. After waking, these individuals took a 30-min break to recover from sleep inertia. Participants in the
wake group watched nature documentaries throughout this 120-min period. In session two, all participants repeated the word-location test (T2) and completed a word pair test.

3. Results

3.1. Pre-study sleep and alertness

Participant groups were matched for hours slept during the night preceding the study [mean ± SEM, sleep group: 7.73 ± .14, wake group: 7.51 ± .15, t(58) = 1.11, p = .27] and alertness (Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973) at Session 1 [mean ± SEM, sleep group: 2.83 ± .13, wake group: 2.50 ± .16, t(58) = 1.65, p = .11] and Session 2 [mean ± SEM, sleep group: 3.30 ± .13, wake group: 3.53 ± .23, t(58) = .88, p = .39].

3.2. Spatial memory and RID

To assess spatial memory accuracy, word-location error scores were created for T1 and T2 by calculating for each word the distance (cm) between the recalled and studied locations (see Table 1, Fig. 1B). RID was indicated when the location recalled at T2 was closer in distance to the location recalled at the immediate test (T1) than to the studied location (i.e., RID index > 0).

An alternative interpretation of a RID index > 0 is that the location memories formed during training were already inaccurate, with T1 and T2 retrievals both being attempts to retrieve the same encoded location. For example, the encoded location of an item might be 2–3 cm away from the studied location and still be classed as correct based on the learning criterion (< 4.8 cm). If T1 and T2 retrieval attempts are both within 1 cm of this inaccurately encoded location, then the T2 location would be closer to the T1 location than to the studied (correct) location. To assess this interpretation, we calculated the distance between the T2 location and the last location recalled by participants in the active learning phase of training (a reasonable indicator of the encoded location, see Table 1). Across all participants, this distance was significantly greater than the distance between T2 and T1 [t(59) = 5.76, p < .001], suggesting that a RID index > 0 reflected a genuine distortion effect of T1 retrieval, rather than poor initial encoding. Like the RID index, this subsidiary measure of distortion [(T2-Active Learning distance) / (T2-T1 distance)] was comparable between the sleep and wake groups [mean ± SEM, sleep group: .42 ± .09, wake group: .31 ± .09, t(58) = .84, p = .41]. See the Supplementary Materials for further analyses utilising active learning locations.

3.3. The effects of sleep

Equivalent RID indices between the sleep and wake groups could mean that sleep had no influence on the retention of distorted memories. However, the results instead suggest that sleep had a positive influence on memory for both the studied locations and the T1-retrieved locations. A Test (T1/T2) × Group (Sleep/Wake) mixed ANOVA conducted on word-location error scores found no main effect of Group [F(1,58) = 2.88, p = .10], but a main effect of Test [F(1,58) = 42.06, p < .001], with studied
locations generally better remembered at T1 than T2. Crucially, there was also a significant interaction \( t(58) = 6.73, p = .01 \): no group difference was observed at T1 \( t(58) = .96, p = .34 \) but the sleep group outperformed the wake group at T2 \( t(58) = 2.27, p = .03 \). Thus, memory decay for studied locations was lower after sleep than wakefulness (see Fig. 2A). Furthermore, T2 locations were closer to the corresponding T1 locations in the sleep group than in the wake group \( t(58) = 2.26, p = .028 \, \text{(Fig. 2B)} \). The effects of sleep on memory for the studied and T1-retrieved locations therefore cancel out to leave the RID index unchanged across the two groups.

A more parsimonious account of these results is that participants in the wake group simply forgot more of the word locations than participants in the sleep group, with locations selected randomly in such cases of forgetting. In this scenario, the error introduced by increased guessing in the wake group at T2 would, relative to the sleep group, increase the distance between T2 and T1. Thus, memory decay for studied locations was significantly lower in the sleep group than the wake group. However, there is a potential problem with this conclusion in that our measures are not fully independent. In particular, when T2-location error is greater than T1-location error (which is on average the case) any variable that draws T2 closer to the studied location will tend to also draw T2 closer to T1. Thus, it could be that sleep is only influencing the ability to retain the studied location and

Table 1 – [A, B] Word-location error (i.e., distance between the recalled locations and studied locations) at T1 or T2. [C] Word-location memory decay (i.e., the increase in word-location error from T1 to T2). [D] Distance between T2 and T1 recalled locations. [E, F] Distance between locations recalled at T1 or T2 and the last locations recalled in the active learning phase of training. [G] RID index [i.e., T2 word-location error – (T1-T2 distance)]. Data are shown in cm (mean ± SEM).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sleep group</td>
<td>3.30 (± .21)</td>
<td>3.64 (± .21)</td>
<td>.33 (± .09)</td>
<td>2.97 (± .21)</td>
<td>3.11 (± .20)</td>
<td>3.39 (± .21)</td>
<td>.67 (± .09)</td>
</tr>
<tr>
<td>Wake group</td>
<td>3.59 (± .22)</td>
<td>4.37 (± .24)</td>
<td>.77 (± .14)</td>
<td>3.74 (± .27)</td>
<td>3.33 (± .23)</td>
<td>4.05 (± .23)</td>
<td>.63 (± .11)</td>
</tr>
</tbody>
</table>

It is tempting therefore to conclude that both original and distorted memory content was better preserved in the sleep group than in the wake group. However, there is a potential problem with this conclusion in that our measures are not fully independent. In particular, when T2-location error is greater than T1-location error (which is on average the case) any variable that draws T2 closer to the studied location will tend to also draw T2 closer to T1. Thus, it could be that sleep is only influencing the ability to retain the studied location
that the effect found for T2-T1 distance is a consequence of this lack of independence. We therefore carried out an additional analysis to demonstrate that there were independent and equivalent benefits of sleep for the studied locations and the locations retrieved at T1.

The three locations of any word box in our experiment — studied (S), T1 and T2 — can be thought of as defining a triangle in the space of the computer screen (see Fig. 1B), with these triangles spread out across the full space depending on the locations of the three components. We aligned these triangles on a trial-by-trial basis to a common co-ordinate system (see Fig. 3). The triangles were translated so that S was at the origin for all trials. They were then rotated about the origin so that the line from S to T1 was aligned to the y-axis, and the triangle was rescaled so that this line had a length of 1 cm for all trials. This aligned the triangles such that T1 occupied the position (0,1) for all trials. The x-axis on this co-ordinate system represented the distance from T2 to the S-T1 line (triangles were flipped if necessary to ensure that all these distances were positive).

The value of this alignment is that it produces T2 locations in a two-dimensional orthogonal space for which the y dimension represents the position of T2 along the line between S and T1. T2 locations can then be assessed in terms of the relative “draw” of both the S and T1 locations for the sleep and wake groups separately. As such, if sleep preferentially consolidates studied or T1-retrieved locations then one would expect sleep participants to have respectively lower or higher T2 y-values than wake participants. Equivalent consolidation of studied and T1-retrieved locations in sleep, by contrast, would predict no group difference. The T2 y-values showed no difference between the sleep and wake groups [mean ± SEM, sleep: .81 ± .07; wake: .79 ± .07; t(58) = .268, p = .79].

Crucially, the x-values for T2 in this space represent the combined draw of S and T1. Here, there was a clear group difference [mean ± SEM, sleep: 1.58 ± .12; wake: 2.08 ± .16; t(58) = 2.47, p = .016]. In sum, this subsidiary analysis supports the conclusion of the main analyses: T2-retrieved locations were influenced by stronger memories for both studied and T1-retrieved locations in the sleep group compared with the wake group.

4. Discussion

We examined the effects of sleep on the consolidation of word-location memories following RID. Three possible outcomes were envisaged, with sleep benefitting only the distorted (T1) locations, sleep benefitting both the studied and the distorted locations (S + T1) or sleep preferentially benefiting the original locations (S). Our data support the second possibility and provide novel evidence that sleep strengthens both original and distorted memory traces following RID.

These findings are in line with the MTT model of memory (Moscovitch & Nadel, 1998; Nadel & Moscovitch, 1997; Winocur et al., 2010, 2013), which proposes that a new hippocampal trace is encoded every time an episodic memory is retrieved. These multiple hippocampal traces are thought to underpin an episodic memory for as long as it is available, and support the development of a neocortical representation that captures the gist of the original experience. From this perspective, erroneously recalled information may be stored as a new hippocampal trace during retrieval, causing memory distortion. Over time and numerous bouts of retrieval, gist-like representations may become increasingly inaccurate as the neocortex draws upon many of these distorted hippocampal memory traces. It should be noted, however, that the present findings have no bearing on questions concerning the continued relevance of hippocampal networks for episodic memory retrieval.

This study provides the first evidence that sleep has independent and equivalent benefits for original and distorted memory content. In the context of an efficient memory system that continually updates and evolves, it can be optimal to use an offline consolidation process that strengthens multiple memory traces rather than just specific information (Dudai, 2012; Schacter et al., 2011). Consolidating only the original memory content would maintain precision but prevent adaptive updating. On the other hand, strengthening only updated memory content following retrieval would lead to an escalation of the distortion that retrieval can introduce. Instead, a consolidation process that strengthens both an original and updated memory trace would lead to both traces contributing to the subsequent retrieval process, with only a relatively small cost in precision.

This account of our data suggests that memory stabilisation in sleep facilitates the integration of original and distorted memory traces during subsequent retrieval. However, we cannot rule out the possibility that our findings, to some extent at least, may also reflect an influence of memory integration during sleep itself. Indeed, many studies have suggested that memory integration is enhanced by sleep (Dumay & Gaskell, 2007; Tamminen, Lambon-Ralph, & Lewis, 2013; Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010), raising the possibility that sleep in the current study supported the amalgamation of original and distorted memory traces into a composite representation. In this scenario, locations retrieved after sleep (vs wake) would retain more features of both the studied and previously retrieved (T1) locations, and, as we observed, would be placed closer in distance to each. Teasing apart the relative contributions of sleep-dependent memory stabilisation and integration to adaptive memory updating will be an important challenge for future research.

Our data are pertinent to understanding the temporal dynamics of memory updating. In earlier work, Bridge and Paller (2012) observed RID following a 24-h delay between the training phase and retrieval phase in which distortion took place. Their findings are reminiscent of reconsolidation, where remote memories are labilised as a result of reactivation and become prone to disruption by interfering information (Nader & Hardt, 2009; Tronson & Taylor, 2007). Central to the reconsolidation hypothesis, however, is that memories have already undergone a significant initial period of consolidation before being reactivated and destabilised, which often occurs over ~24 h in human studies (Porcato et al., 2007; Hupbach, Gomez, Hardt, & Nadel, 2007). In the current study, RID was observed when the first retrieval phase took place immediately after encoding, meaning that memory stabilisation had not occurred. Memory updating via retrieval alone therefore appears to unfold in the absence of any significant delay between training and test.

In summary, retrieval promotes memory updating via the storage of retrieved information, sometimes causing
distortion (Bridge & Paller, 2012). Our findings suggest that RID leads to the formation of an additional memory trace that corresponds to a distorted variant of the same encoding event. Moreover, our data suggest that sleep strengthens original and distorted memory traces to similar extents, enhancing access to both during subsequent retrieval operations.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Acknowledgements

This work was supported by an Economic and Social Research Council (ESRC) grant (ES/I038586/1) awarded to MGG and SL. The ESRC had no involvement in the conduct of this research or the decision to submit the article for publication. We are grateful to Donna Bridge and Emma James for critical reading of the manuscript. We also thank two anonymous reviewers for their very thoughtful and insightful comments.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.cortex.2017.10.005.

REFERENCES


