In press at Cortex

Sleep not just protects memories against forgetting, it also makes them more accessible

Nicolas Dumay

Department of Psychology, University of Exeter, United Kingdom, and BCBL (Basque Center on Cognition, Brain and Language), Spain

Word count:

Abstract – 199

Main text, plus footnotes – 5,121

Correspondence:

Nicolas Dumay

Department of Psychology

University of Exeter

Perry Road

Exeter

EX4 4QG

United Kingdom

E-mail: nicolas.dumay@gmail.com

Phone: +44 1392 724666

Abstract

Two published datasets (Dumay & Gaskell, 2007, *Psychological Science*; Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010, *Journal of Neuroscience*) showing a positive influence of sleep on declarative memory were re-analyzed, focusing on the "fate" of each item at the 0-hr test and 12-hr retest. In particular, I looked at which items were retrieved at test, and "maintained" (i.e., not forgotten) at retest, and which items were not retrieved at test, but eventually "gained" at retest. This gave me separate estimates of protection against loss and memory enhancement, which the classic approach relying on net recall/recognition levels has remained blind to. In both free recall and recognition, the likelihood of maintaining an item between test and retest, like that of gaining one at retest, was higher when the retention interval was filled with nocturnal sleep, as opposed to day-time (active) wakefulness. And, in both cases, the effect of sleep was stronger on gained than maintained items. Thus, if sleep indeed protects against retroactive, unspecific interference, it also clearly promotes access to those memories initially too weak to be retrieved. These findings call for an integrated approach including both passive (cell-level) and active (systems-level) consolidation, possibly unfolding in an opportunistic fashion.

Keywords: memory consolidation, sleep, forgetting, reminiscence, item fate

"The analysis of recall patterns has clearly shown that the "average" item is a highly abstract and elusive entity having no readily identifiable counterparts in the empirical realm." (Tulving, 1967, p. 183).

1. Introduction

In a study titled "Obliviscence during sleep and waking" Jenkins and Dallenbach (1924) had two of their colleagues learn lists of nonsense syllables either in the morning or late evening. By testing recall after intervals ranging between one and eight hours, these researchers found that the presence of sleep in the retention interval had a protective influence: forgetting curves were less steep for intervals filled with sleep than for those filled with active wake.

Nine decades down the forgetting curves, the beneficial impact of sleep on memory is well established (see Wixted & Cai, 2014, for a review). We know that the sooner the learner sleeps after encoding, the better the memory retention (Ekstrand, 1972; Gais, Lucas, & Born, 2006; Payne et al., 2012). We also know that which sleep component is key to the consolidation process depends on the type of knowledge to be consolidated. For instance, the conscious recollection of factual information or of our past experience (i.e., "declarative" memory) almost exclusively benefits from slow-wave (slow oscillations) sleep (Plihal & Born, 1997, 1999; Yaroush, Sullivan, & Ekstrand, 1971). By contrast, the positive effect of sleep on newly learnt perceptual or motor skills, i.e., procedural memory, shows a more complex picture, in which REM (rapid-eye movement) and non-REM sleep brain correlates have been implicated, sometimes in synergy (e.g., Gais, Plihal, Wagner, & Born, 2000; Gaskell et al., 2014; Karni, Tanne, Rubenstein, Askenasy, & Sagi, 1994; Rasch, Gais, & Born, 2009; Stickgold, James, & Hobson, 2000).

One of the current debates, however, is whether sleep makes declarative (hippocampus-dependent) memories more vivid, and thus more accessible, than they were just after encoding, or if instead sleep merely protects them from the deleterious effect of retroactive interference. According to the "active consolidation" account (Diekelmann & Born, 2010), sleep is key to memory consolidation because slow-wave sleep promotes "neural replay" (Wilson & McNaughton, 1994). In slow-wave sleep, new memories encoded in the hippocampus are repeatedly re-activated, which drives concurrent reactivation of cortical regions implicated in their initial capture (Ji & Wilson, 2007). By this feedback from the hippocampus to the cortex, or "systems consolidation", newly acquired memories are effectively re-experienced, with the result that cortical representations are strengthened. They are also better integrated in pre-existing cortical networks, because replay also reactivates similar, long-consolidated material. Hence, according to this account, which emphasizes the role of the hippocampus as the sparring partner of the neocortex, sleep has the potential to make memories more accessible.

In contrast, theories grounded in the notion of retroactive interference (Wixted, 2004; see also Mednick, Cai, Shuman, Anagnostaras, & Wixted, 2011) insist that during slow-wave sleep synaptic plasticity in the hippocampus is null. Consequently, hippocampal resources that would otherwise be allocated to new encoding can now be used to consolidate, at the cell level, memories formed prior to (slow-wave) sleep. Amongst these, those formed earlier during the wake should be more eroded. Thus, memory consolidation is fundamentally the antidote to retroactive, unspecific interference and thus reduces its product: forgetting. Any factor that reduces the encoding activities of the hippocampus ipso facto promotes consolidation. As this account assumes that hippocampal consolidation triggers systems consolidation, anything that is forgotten by the hippocampus cannot be recovered via neural replay (Wixted & Cai, 2014, p.

30). Thus, according to the anti-forgetting view, sleep—like any other interval of reduced encoding—can at best stabilize newly formed declarative memories; it cannot make them more accessible.

So far, the "active consolidation" account has received strong empirical support from demonstrations that declarative memories are better preserved if, while in slow-wave sleep, participants are cued by an odor or sound also present during encoding (Rash, Buchel, Gais, & Born, 2007), or by translation equivalents, in the case of word lists (Schreiner & Rasch, 2014). Frustratingly, however, the data supposedly speaking to the issue of sleep-dependent trace enhancement remain ambiguous, providing little support for the idea.

I suggest that this state of affairs is due to the fact that researchers have been relying exclusively on net performance. However, as Tulving (1964, 1967), amongst others, pointed out, this approach is inherently blind to fluctuations at the item-level. Consequently, even though the typically observed pattern is that the sleep group simply shows less forgetting than the wake group, it may well be that a sleep-dependent trace enhancing mechanism is actually counteracting the effect of a task-specific component that worsens performance at retest. Conversely, and by the same logic, finding that the sleep group shows more improvement at retest (compared to the initial test) than the wake group does not necessarily provide evidence for sleep-dependent trace enhancement. A sleep-dependent anti-forgetting mechanism could just supplement a task-specific component which, irrespective of group, helps to maintain/improve performance. Therefore, without information on the trajectory of each item between test and retest, it is misguided to use declarative tasks and make inferences about consolidation mechanisms supposedly acting on individual representations.

Consequently, in the present research I tracked the "fate" of individual items from an initial test to a post-sleep retest to better assess the impact of sleep on memory. Specifically, I distinguished between items that were retrieved at test and "maintained" (i.e., not forgotten) at retest, and items that were not retrieved at test, but were eventually recovered (i.e., "gained") at retest—a phenomenon typically referred to as "reminiscence". This gave me separate estimates of protection against loss and of memory enhancement. These provide a means to determine whether sleep only boosts protection against loss, which would support the anti-forgetting account (Wixted, 2004), or whether it also boosts memory accessibility, which would support the active consolidation account (Diekelmann & Born, 2010).

Item-fate analysis is routinely performed in cognitive psychology (Ballard, 1913; Brown, 1923; Erdelyi, 1984; Tulving, 1964, 1967, just to cite the pioneers). However, it is only recently that Fenn and Hambrick (2013) used this approach to look at the effect of sleep, unfortunately rather unconvincingly, as I will demonstrate. Fenn and Hambrick had 354 participants learn pairs of semantically related items (e.g., table-chair) either in the evening or in the morning. After testing recall of the second member of each pair (e.g., table-?) immediately and after twelve hours, and finding improvement for both groups, they classified responses following the above maintained vs. gained distinction. As sleep had its strongest impact on maintained items (+1.35 item compared to wake; versus +.44 for gained), the authors concluded that "[…] loss prevention may primarily account for the effect of sleep on declarative memory consolidation" (p. 501).

As can be seen on their Fig. 1, however, performance was dangerously close to ceiling, especially at the 12-hr retest (on average, 35 items recalled out of 40). To reassure the reader, Fenn and Hambrick crosschecked their results, excluding all participants with a retest score four items (i.e., the average improvement across the two groups), or less, away from the ceiling (see their Footnote 4 and Supplemental Materials). This is dubious: the average improvement is itself determined by ceiling height, and therefore, it cannot be used as a safeguard.

To assess whether this procedure was at all effective, I computed the correlation between the 0-hr score and the magnitude of the 12-hr improvement for the 322 remaining participants. As shown Fig. 1, there was still a negative correlation: the better a participant was at test, the less (s)he improved at retest. In other words, the restricted dataset was still constrained by the ceiling. Clearly, with so little room for improvement, this experiment hardly gave sleep a chance to demonstrate its potential for enhancing memory—the same criticism applies to a follow-up just published (Fenn & Hambrick, 2015), which used the same parameters, task, crosscheck, and another 435 participants.

Given this concern, I re-analyzed the data from two published studies (Dumay & Gaskell, 2007; Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010; henceforth 'DG' and 'TPSWG'), that showed a positive effect of sleep on free recall under conditions in which performance was not constrained by the ceiling (or the floor). These studies used similar methodologies, which makes them easy to combine. Participants were trained on made-up words either in the evening or in the morning in a phoneme detection task, and urged to learn. Memory tests were then administered immediately and again after twelve hours, meaning, after nocturnal sleep or a day awake, respectively. Crucially, because of the nature of the to-be-learnt stimuli, all competitors of existing words (e.g., 'caravoth' for 'caravan'), both studies could demonstrate overnight integration of the new items in lexico-semantic memory (cf. Gaskell & Dumay, 2003; Leach & Samuel, 2007). In both studies, emergence of the competitive influence from the novel words on perception of their base words was contingent upon the occurrence of sleep during the test-retest interval (see Dumay & Gaskell, 2005). Thus, on the basis of these non-linear changes, we know that for the sleep group something akin to systems consolidation had been triggered (e.g., Takashima et al., 2009).

With this appropriate data set, item fate can show whether the net improvement in declarative knowledge observed after sleep was driven by protection against loss, by memory enhancement, or by both. If protection against forgetting is the driving force, the sleep group should have a higher proportion than the wake group only on maintained items; in contrast, if active trace enhancement is operating, then the sleep group should instead differ from the wake group only on gained items.

2. Original Methods

In DG, training involved 24 made-up words (such as 'frenzilk') created by adding two consonants to vowel-final nouns, each played 36 times (see Dumay & Gaskell, 2012, for the list of materials). In TPSWG, the novel words were 30 onset-overlapping competitors (such as 'caravoth'); each was played 30 times in phoneme detection, plus four times in oral repetition. Training took place between 8:00 and 8:45 a.m./p.m. in DG, and between 9:15 and 10 p.m., and 8:30 and 10 a.m. in TPSWG. Retest followed exactly 12.5 hr after the immediate test in DG, whereas this was after 9.5 hr for the wake group and 10 hr for the sleep group in TPSWG.[i] In both studies, free recall always came first in the series of declarative tasks. However, as the focus was on whether the newly learnt, novel word representation influenced perception of the long-consolidated base word, priority was given to the so-called "lexical competition" test, in which only the overlapping base words (e.g., 'caravan'), not the made-up words themselves, were presented. In the free recall task, participants had three minutes to report as many novel words as they could remember from the training phase. Responses were digitally recorded and checked for accuracy afterwards. This was followed by a two-alternative recognition test only for half the participants in DG, and a cued recall and an old/new recognition test in TPSWG. See the original articles for further details.

3. Analysis One—Free Recall

The first analysis combined the free recall data of the 42 "best" participants out of the 64 tested by DG and of the 30 "best" participants out of the 59 reported in TPSWG (see below). The resulting sleep and wake groups thus comprised 36 participants each, of whom 21 came from DG. The use of a restricted sample was motivated by the presence of a floor effect in the wake group in the total sample, as shown by a significant negative correlation between test scores and amount of forgetting at retest ($r(61) = -.30$, $p < .02$). To eliminate this problem, participants with the poorest test scores were removed until this correlation became non-significant, and close to zero ($r(36) = -.08$, $p > .65$). The sleep group showed no sign of a ceiling/floor effect in the total sample ($r(62) = -.09$, $p > .50$), or in the restricted set ($r(36) = -.06$, $p > .73$). Using this procedure assured that in *both* groups performance was free to vary either way between the 0-hr test and the 12-hr retest. Note that the full dataset showed overall the same pattern of results as that reported for the restricted sample.

I first re-analyzed the data using the classical approach based on percent correct. Included in the analysis of variance (ANOVA) was a dummy factor coding which study the participant came from. As this factor had no interactive effect, I will not consider it any further.

As shown Fig. 2a, the sleep and wake groups did not differ significantly on their performance at the 0-hr test (21.4 vs. 20.7%; $F < 1$), but diverged substantially by the 12-hr retest (Group x Session: $F(1,68) = 24.67$, $p < .0001$). Although the sleep group simply reproduced the improvement (+5.7%) already seen in the individual datasets ($F(1,34) = 13.21$, $p < .001$), the wake group, now unbound, statistically confirmed the trend towards forgetting only suggested in the original studies (-4.8%; *F*(1,34) = 11.47, *p* < .002). Similar ANOVAs on arcsined proportions showed exactly the same results.

3.1 "Item fate" analysis

To examine the effect of sleep on protection against loss and on memory enhancement separately, I first classified each item on its recall trajectory between test and retest for each participant. Each item was entered in one of the following categories: (1) "never recalled"; (2) "lost" (i.e., recalled only at test); (3) "maintained" (i.e., recalled at both test and retest); or (4) "gained" (i.e., recalled only at retest). As "never recalled" items were not useful to my purpose and "lost" items were just the complement to maintained items, the first two categories were discarded.

The next step was to derive measures that took into account how many items can theoretically be maintained, and how many items can theoretically be gained, between test and retest. Note that by just relying on raw item counts, Fenn and Hambrick (2013, 2015) failed to take these dimensions into account.[ii,iii] Consequently, for each participant I computed proportions, by taking, respectively,

(1) for the maintained items: the number of items maintained at retest relative to the number of items recalled at test (i.e., the maximum number of items s/he could have maintained);

(2) for the gained items: the number of items gained at retest relative to the total number items minus the number of items recalled at test (i.e., the maximum number of items s/he could have gained). This is exactly the procedure introduced by Roediger and Challis (1989, p. 179).

Gaining an item (i.e., "reminiscence") was much less likely than maintaining one, as shown by the proportions average across conditions (gained: 12.1% vs. maintained: 57.0%). Crucially, however, gained items were the ones that benefitted most from the participants sleeping during the test-retest interval, as indicated by their sleep-to-wake ratio of 2.30 (from 7.3% for wake, to 16.9% for sleep) compared to 1.49 for maintained items (from 45.6% to 68.3%, respectively).

To back this up and test for the interaction, I first rescaled the proportions of maintained and gained to take into account the difference in probability in the two types of recall trajectory. A normalization-by-the-mean was applied, in which the proportion of maintained and the proportion of gained were divided by their respective means (by participants), computed over the whole dataset in order to be neutral with respect to the sleep/wake contrast (see Fig. 2b). I then ran an ANOVA similar to the one performed on percent correct. The effect of sleep was significant for both gained items (from .61 for wake, to 1.39 for sleep; $F(1,68) = 15.79$, $p < .0002$) and maintained items (from .80 to 1.20, respectively; $F(1,68) = 11.50$, $p < .002$), but the interaction was significant as well ($F(1,68) = 4.07$, $p < .05$), which confirmed that sleep had a stronger influence on gained items.[iv]

In view of these results, clearly, sleep protects against forgetting, as shown by the higher proportion of maintained items after sleep, compared to wake. But its most compelling impact is on promoting access to memories initially too weak to be recalled, as shown by its effect on gained items. To provide a conceptual replication of these findings, in a different declarative task, I applied the same analysis to the old/new speeded recognition data from the 60 participants tested by TPSWG. Note that in the original paper only the latency data are reported.

4. Analysis Two—Old/New Recognition

In TPSWG the recognition task included only half of the 30 target made-up words at the 0-hr test (mixed with minimally diverging foils, e.g., "caravol"), but all of them at the 12-hr retest. On each trial participants had to decide, as quickly as possible, whether they recognized the stimulus as part of the to-be-learnt set. The present analysis is based on these 15 items (which varied across participants).

Although the wake group exhibited no significant variation in recognition speed between the test and the retest (see their Fig. 2c), the accuracy data—as re-analyzed here—showed forgetting (-12.2%; $F(1,29) = 13.28$, $p < .002$; see this Fig. 2c). In contrast, the sleep group showed the reverse: despite a robust improvement in terms of speed (possibly due to consolidated practice, which the authors did not allude to), these participants showed no reliable change in accuracy (+ .4%; $F < 1$; Group x Session: $F(1,58) = 7.37$, $p < .009$). Again, ANOVAs on arcsined proportions showed exactly the same results.

Hence, on the surface (i.e., based on net performance), the recognition data seem to support the view that sleep only prevents memory loss. To make this more concrete, just compare Fig. 2a with Fig. 2c. Given this apparent conflict with the free recall results, an analysis of item trajectory/fate is clearly needed.

4.1 "Item fate" analysis

As in free recall, the probability of recovering a previously unretrieved item was weaker than that of maintaining one, across test and retest (i.e., gained: 47.3% vs. maintained: 80.7%). But once again, gained items had a higher sleep-to-wake ratio, of 1.60 (from 36.3% for wake, to 58.2% for sleep) than maintained items, with 1.06 (from 78.3% to 83.0%; see Fig. 2d), indicating that gained items were the ones to benefit most from sleep.

This difference was attested by the presence of a significant interaction between group and item trajectory ($F(1,58) = 6.99$, $p < .02$) on the normalized proportions. In contrast to free recall, however, the effect of sleep was significant only for the gained items ($F(1,58) = 8.47$, $p < .006$), not for the maintained items ($F(1,58) = 1.28$, $p > .26$).[v] In short, although the maintained items showed a weaker sleep effect here, compared to free recall, the same configuration emerged: sleep did not just prevent forgetting, it also enhanced memory accessibility. In fact, the memory enhancement was stronger than the prevention of forgetting.

5. Correlational Analysis

To determine whether this accessibility boost was driven by the same mechanism as protection against loss, I computed the correlations between participants' ability to maintain items and their ability to gain others during the twelve hours of the test/retest interval. The two tasks (free recall and old/new recognition) showed the same dissociation (see Fig. 3 a-d): whereas the proportions of maintained and gained items were positively correlated after wake ($r(36) = .42$, $p < .02$; $r(30) = .38$, $p < .04$), they were not correlated after sleep ($r(36) = .19$, $p > .27$; $r(30) = .10$, $p > .6$, respectively). The positive correlations after wake may indicate that participants who were good at structuring their recall, not only maintained more items, but also had more time left to look again and gain items. But this is as far as it goes: these differences in recall strategies did not seem determine how well participants maintained and/or gained items in the recognition task, as indicated by the absence of correlations across tasks (maintained: $r(29) = .16$, $p > .42$; gained: $r(29) = .17$, $p > .37$; pooled: $r(29) = .05$, $p > .80$)[vi]. More strikingly, the presence of a non-linear change, from covariation (during wake) to independence (during sleep), tells us that during sleep something different was going on, and that protection against loss and accessibility enhancement, at that point, no longer reflected the same mechanism.

6. Discussion

Sleep improves the accessibility of declarative memories more than it protects them against interference. In two studies in which net recall/recognition showed forgetting after 12 hours of wake, item fate analyses revealed that the post-sleep advantage was predominantly due to sleep promoting access to memory traces that had initially been too weak to be retrieved. This is at odds with the conclusions drawn by Fenn and Hambrick (2013), who argued only for protection against loss. As I have shown, their dataset was flawed by a ceiling effect, which may have masked the true effect of sleep on memory. This also casts doubt on their recent paper (Fenn & Hambrick, 2015) on consolidation and IQ, in which in all likelihood a similar flaw was at play.

The observed post-sleep boost in memory accessibility is not easily explained by the "anti-forgetting" account (Wixted, 2004; see also Mednick et al., 2011). It is even less so in view of the fact that sleep seems to wipe out the correlations between the proportion of maintained and gained items that can be seen after wake (see above). In this model, the fundamental hypothesis is that, during slow-wave sleep, the hippocampus is unable to form new memories. As new hippocampal encoding would necessarily erode recently formed traces, by blocking this process sleep could only prevent retroactive interference, and thereby memory loss (cf. Wixted, 2004, p. 261).

Still, proponents of the anti-forgetting view may argue that the sleep-associated supplement of gained items is tied to the endogenous nature of the free recall task. Attempts made at recalling items at the 0-hr test could set the brain in motion to search actively—even offline—for those memories that it failed to access. Under this view, the post-sleep accessibility boost therefore would not reflect a genuine representational enhancement, but rather the fact that this offline search has been promoted by sleep, as is the case for similar phenomena, such as insight (Fischer, Drosopoulos, Tsen, & Born, 2006) and problem solving (Gupta, Jang, Mednick, & Huber, 2012; Sio, Monaghan, & Ormerod, 2013). And since participants in the recognition task were all tested in free recall as well, this explanation would hold also for the boost observed in the more exogenous, old/new recognition.

To test this hypothesis, I looked at the trajectory of each item in old/new recognition, taking into account whether or not the item had been successfully recalled at the 0-hr test. Since the putative search should not apply to items correctly recalled at test, these should return a weaker accessibility boost (compared to unrecalled items) —if not also less protection against loss—at the post-sleep recognition retest. As shown by the pooled proportions of gained and maintained items (taken as an estimate of overall consolidation), the sleep advantage in the recognition data (expressed as the sleep-to-wake ratio) was unaffected by the fate of the item at the 0-hr recall (recalled: 1.27 (from 1.12 to .88) vs. unrecalled: 1.21 (from 1.10 to .90); $F(1,58) = 1.35$, $p > .25$). This clearly does not support the "offline search" interpretation.

The post-sleep boost in memory accessibility seems to indicate instead that declarative memories can be sharpened overnight, an option clearly envisaged by the active consolidation account (Diekelmann & Born, 2010). If by reactivating cortical regions that originally contributed to its encoding of a new trace, the hippocampus promotes binding amongst those regions, changes in memory strength could ensue, resulting in more accessible, more vivid representations. From this perspective, the present results nicely complement the cued reactivation effects obtained in slow-wave sleep (e.g., Rasch & Born, 2007) in supporting this account. They also predict that the latter effects should be most visible on gained (as opposed to maintained) items.

The benefit of sleep on maintained items obtained in the free recall task is, nonetheless, in keeping with the predictions of the anti-forgetting account (Wixted, 2004). If the impossibility for the hippocampus to encode during slow-wave sleep means that memories formed shortly beforehand can be hardened more promptly and with minimal damage, then sleep (as opposed to wake) should result in less forgetting. This is exactly what the present data demonstrate at the item/representational level, at least when the total number of items that could have been maintained is taken into account (cf. Roediger & Challis, 1989).

In sum, a statistical approach that focuses on item fate while bearing in mind what is theoretically possible within each category of item-trajectory provides a much more nuanced and informative picture of the role of sleep on memory consolidation. In this picture, we see distinct evidence confirming complementary aspects of both the anti-forgetting and the active-consolidation account (e.g., Diekelmann & Born, 2010; Wixted, 2004). However, we also see that, contrary to what net performance has led us to believe, protection against forgetting plays only a minor part.

Although the present data compared nocturnal sleep to active wake, they should not be taken to indicate that sleep plays a unique role in consolidating memories. According to the notion of "opportunistic" consolidation (Mednick et al., 2011, Wixted & Cai, 2014; see also Hasselmo, 1999), the brain consolidates recently acquired information whenever the hippocampus is not busy encoding new traces. A period of reduced encoding, whether achieved by slow-wave sleep, or any other means, including wakeful rest (e.g., Dewar, Alber, Butler, Cowan, & Della Sala, 2012) would be sufficient to promote synaptic (hippocampal) consolidation and systems consolidation, possibly with some temporal overlap. Therefore, the sleep manipulations which I relied upon may be best seen as manipulations of the amount of trace encoding between acquisition and retest. In any case, if one assumes that protection against loss and increased accessibility reflect synaptic vs. systems consolidation, the overnight disappearance of their correlation suggests that although they may be triggered by the same factor, synaptic and systems consolidation quickly become free from each other.

Acknowledgments

I thank Kimberly Fenn and Jakke Tamminen, for making their raw data available. I am also grateful to Arty Samuel, Magdalena Abel, Henry Roediger III and two anonymous reviewers, for commenting on earlier versions of this manuscript.

References

Ballard, P. B. (1913). Oblivescence and reminiscence. *British Journal of Psychology Monograph Supplements*, 1, 1-82.

Brown, W. (1923). To what extent is memory measured by a single recall trial? *Journal of Experimental Psychology*, 6, 377-382.

Dewar, M., Alber, J., Butler, C., Cowan, N., & Della Sala, S. (2012). Brief wakeful resting boosts new memories over the long term. *Psychological Science*, 23, 955-960.

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, 11, 114-126.

Dumay, N., & Gaskell, M. G. (2005). Do words go to sleep? Exploring consolidation of spoken forms through direct and indirect measures. *Behavioral and Brain Sciences*, 28, 69-70.

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, 18, 35-39.

Dumay, N., & Gaskell, M. G. (2012). Overnight lexical consolidation revealed by speech segmentation. *Cognition*, 123, 119-132.

Ekstrand, B. R. (1972). To sleep, perchance to dream (about why we forget). In C. P. Duncan, L. Sechrest, and A. W. Melton (Eds.), *Human memory: Festschrift for Benton J. Underwood* (pp. 59-82). New York: Appelton-Century-Crofts.

Erdelyi, M. (1984). The recovery of unconscious (inaccessible) memories: Laboratory studies of hypermnesia. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 18, pp. 95-127). New York: Academic Press.

Fenn, K. M., & Hambrick, D. Z. (2013). What drives sleep-dependent memory consolidation: Greater gain or less loss? *Psychonomic Bulletin & Review*, 20, 501-506.

Fenn, K. M., & Hambrick, D. Z. (2015). General intelligence predicts memory change across sleep. *Psychonomic Bulletin & Review*, 22, 791-799.

Fischer, S., Drosopoulos, S., Tsen, J., & Born, J. (2006). Implicit learning-Explicit knowing: A role for sleep in memory system interaction. *Journal of Cognitive Neuroscience*, 18, 311-319.

Gais, S., Lucas, B., & Born, J. (2006). Sleep after learning aids memory recall. *Learning & Memory*, 13, 259-262.

Gais, S., Plihal, W., Wagner, U., & Born, J. (2000). Early sleep triggers memory for early visual discrimination skills. *Nature Neuroscience*, 3, 1335-1339.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, 89, 105-132.

Gaskell, M. G., Warker, J., Lindsay, S., Frost, R., Guest, J., Snowdon, R., & Stackhouse, A. (2014). Sleep underpins the plasticity of language production. *Psychological Science*, 25, 1457-1465.

Gupta, N. G., Jang, Y., Mednick, S. C., & Huber, D. E. (2012). The road not taken: Creative solutions require avoidance of high-frequency responses. *Psychological Science*, 23, 288-294.

Hasselmo, M. E. (1999). Neuromodulation: acetylcholine and memory consolidation. *Trends in Cognitive Sciences*, 3, 351-359.

Jenkins, J. M., & Dallenbach, K. M. (1924). Obliviscence during sleep and waking. *American Journal of Psychology*, 35, 605-612.

Ji, D., & Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience*, 10, 100-107.

Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J. M., & Sagi, D. (1994). Dependence on REM sleep of overnight perceptual skill. *Science*, 265, 679-682.

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, 55, 306-353.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1, 476-490.

Mednick S. C., Cai D. J., Shuman T., Anagnostaras S., Wixted J. T. (2011). An opportunistic theory of cellular and systems consolidation. *Trends in Neurosciences*, 34, 504-514.

Payne, J. D., Tucker, M. A., Ellenbogen, J. M., Wamsley, E. J.,Walker, M. P., Schacter, D. L., & Stickgold, R. (2012). Memory for semantically related and unrelated declarative information: The benefit of sleep, the cost of wake. *PLoS ONE*, 7, e33079.

Plihal, W., & Born, J. (1997). Effects of early and late nocturnal sleep on declarative and procedural memory. *Journal of Cognitive Neuroscience*, 9, 534-547.

Plihal, W., & Born, J. (1999). Effects of early and late nocturnal sleep on priming and spatial memory. *Psychophysiology*, 36, 571-582.

Rasch, B., Gais, S., & Born, J. (2009). Impaired off-line consolidation of motor memories after combined blockade of cholinergic receptors during REM sleep-rich sleep. *Neuropsychopharmacology*, 34, 1843-1853.

Rasch, B., Buchel, C., Gais, S., & Born, J. (2007). Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science*, 315, 1426-1429.

Roediger, H. L., & Challis, B. H. (1989). Hypermnesia: Improvements in recall with repeated testing. In C. Izawa (Ed.), *Current Issues in Cognitive Processes: The Tulane Floweree Symposium on Cognition* (pp. 175-199). Hillsdale (NJ): Lawrence Erlbaum Associates.

Schreiner, T., & Rasch, B. (2014). Boosting vocabulary learning by verbal cueing during sleep. *Cerebral Cortex* (Epub ahead of print). DOI: 10.1093/cercor/bhu139

Stickgold, R., James, L. & Hobson, J. A. (2000). Visual discrimination learning requires sleep after training. *Nature Neuroscience*, 3, 1237-8.

Sio, U. N., Monaghan, P., & Ormerod, T. (2012). Sleep on it, but only if it is difficult: Effects of sleep on problem solving. *Memory & Cognition*, 41, 159-166.

Takashima, A., Nieuwenhuis, I. L. C., Jensen, O., Talamini, L. M., Rijpkema, M., & Fernández. G. (2009). Shift from hippocampal to neocortical centered retrieval network with consolidation. *Journal of Neuroscience*, 29, 10087-10093.

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *Journal of Neuroscience*, 30, 14356-14360.

Tulving, E. (1964). Intratrial and intertrial retention: Notes towards a theory of free recall verbal learning. *Psychological Review*, 71, 219-237.

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 6, 175-184.

Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, 265, 676-679.

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, 55, 235-269.

Wixted, J. T., & Cai, D. J. (2014). Memory consolidation. In K. Ochsner & S. Kosslyn (Eds.), *Oxford Handbook of Cognitive Neuroscience* (pp. 1-59). New York: Oxford University Press.

Yaroush, R., Sullivan, M. J., & Ekstrand, B. R. (1971). The effect of sleep on memory: II. Differential effect of the first and second half of the night. *Journal of Experimental Psychology*, 88, 361-366.

Figure captions

Fig. 1. Participants' performance at the 0-hr test plotted against magnitude of the 12-hr improvement (%) in Fenn and Hambrick's (2012) restricted dataset, plus trend line and correlation value.

Fig. 2. **(a)** Free recall percent correct for the sleep and wake groups in the DG and TPSWG restricted dataset ($n = 72$). **(b)** Normalized proportions of "maintained" and "gained" items. **(c)** Old/new recognition percent correct in TPSWG ($n = 60$). **(d)** Normalized proportions of "maintained" and "gained" items. Error bars show 95% confidence intervals based on Loftus and Masson (1994).

Fig. 3. Proportion of gained items plotted against that of maintained items, plus trend line and correlation value for: **(a)** free recall after wake, **(b)** old/new recognition after wake, **(c)** free recall after sleep, and **(d)** old/new recognition after sleep. Note. In (a) the correlation is still substantial even after removing the outlier with a normalized proportion gained of 3.75 ($r(35) = .39$, $p < .02$; see alternate plot).

Fig. 1

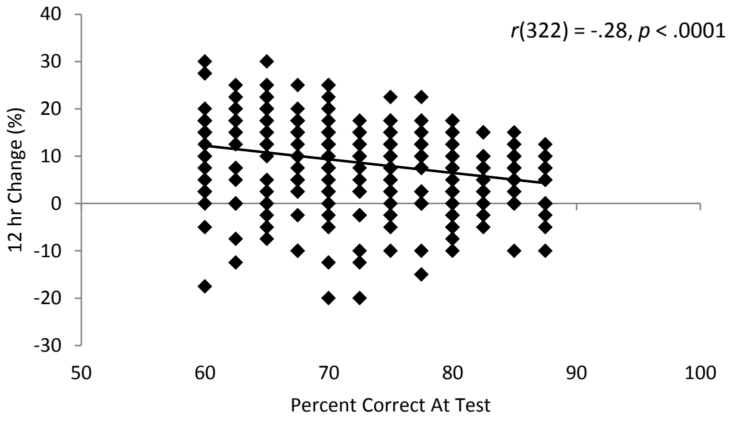

r(322) = -.28, *p* < .0001
12 hr Change (%)
40
30
20
10
0
-10
-20
-30
50
60
70
80
90
100
Percent Correct At Test

Fig. 2

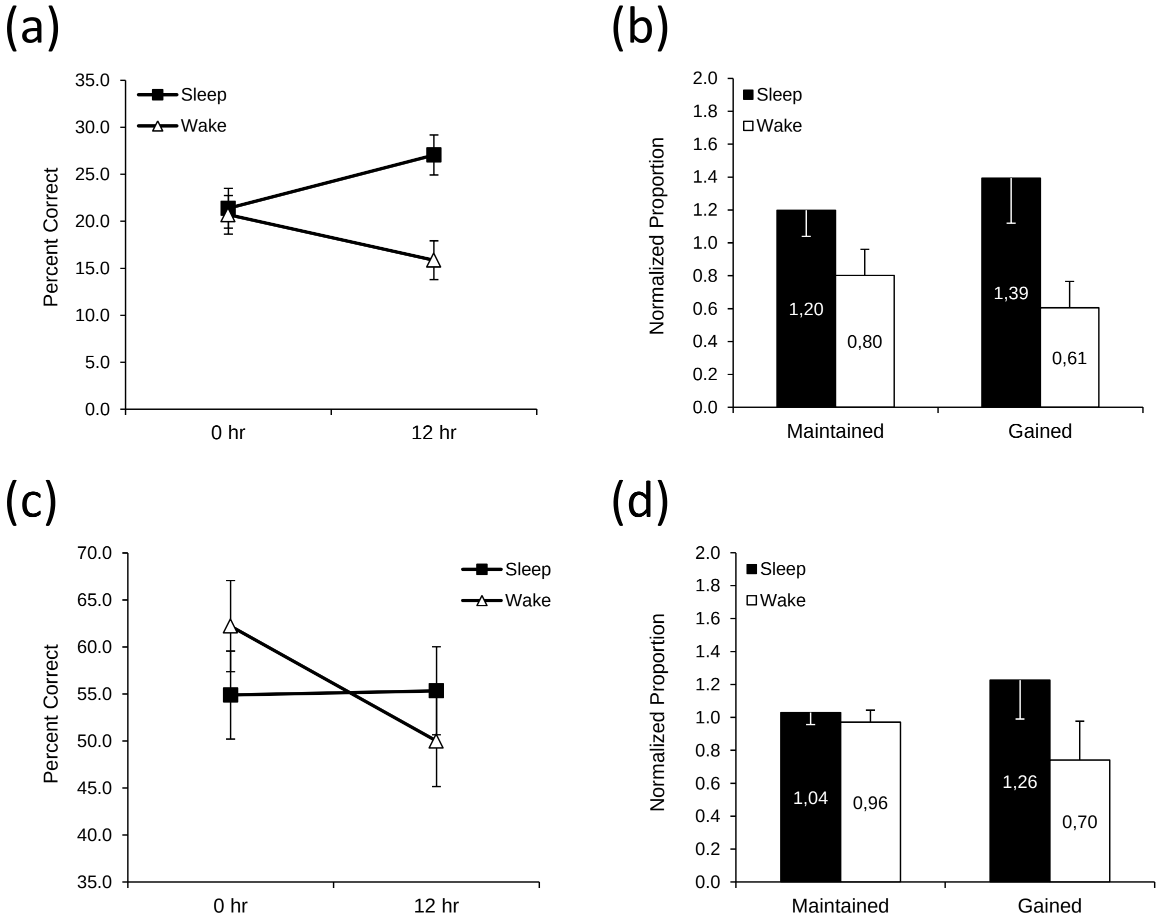
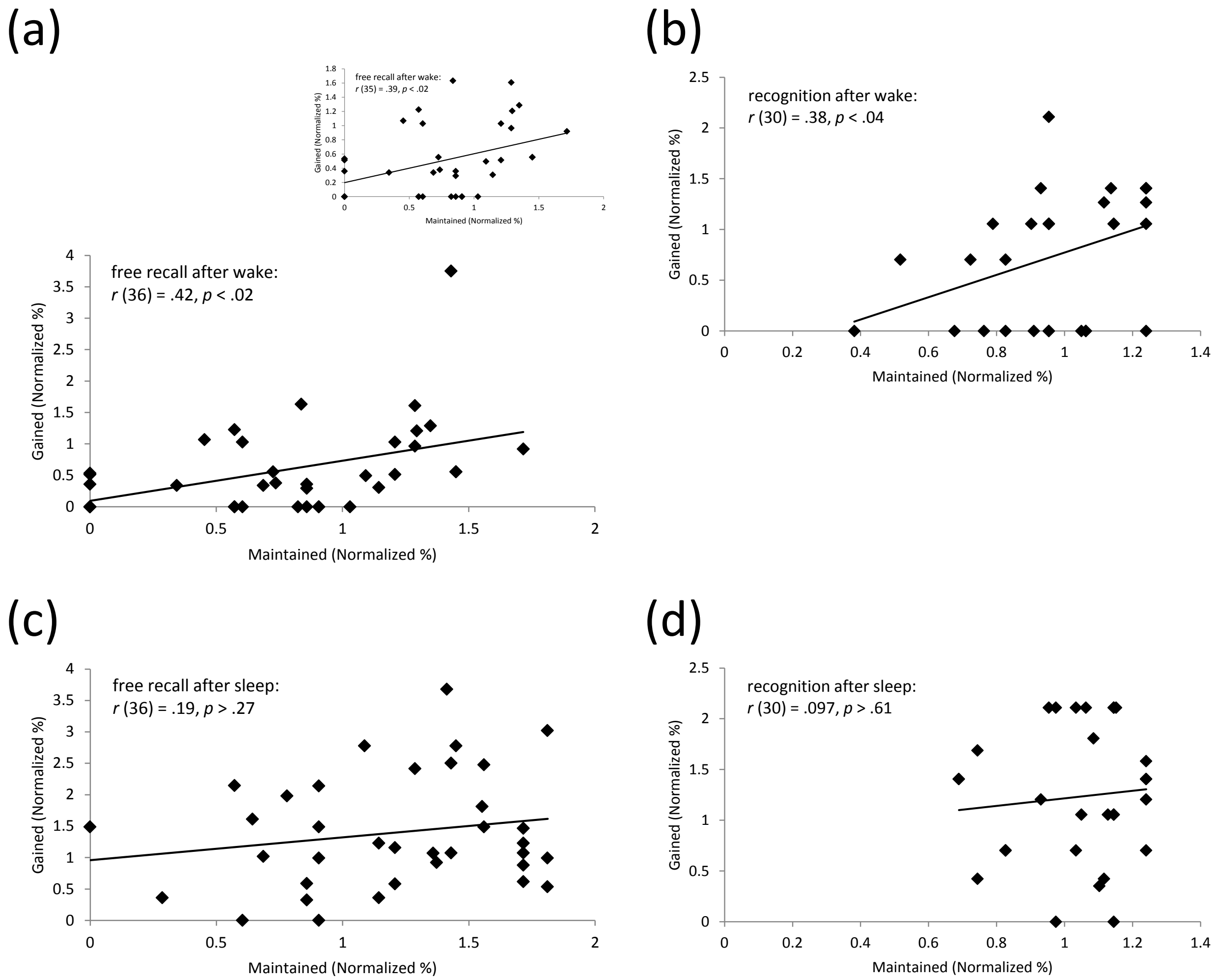

(a)
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
Percent Correct
Sleep
Wake
0 hr
12 hr
(b)
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Normalized Proportion
■ Sleep
□ Wake
1,20
0,80
1,39
0,61
Maintained
Gained
(c)
70.0
65.0
60.0
55.0
50.0
45.0
40.0
35.0
Percent Correct
Sleep
Wake
0 hr
12 hr
(d)
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Normalized Proportion
■ Sleep
□ Wake
1,04
0,96
1,26
0,70
Maintained
Gained

Fig. 3



---

[i] The data of the second retest at 24 hours or seven days are irrelevant to the present argument.

[ii] Just to illustrate, if participant A maintains 4 items at retest out of 8 recalled at test, and participant B also maintains 4 items, but out of 6 recalled at test, Fenn and Hambrick would give them both a count of 4 maintained items, while in fact participant B maintains relatively more of what s/he initially acquired. Similarly, if both participants gain 4 items at retest, participant A

---

gains more of what s/he did not recall at test (say 12 if the maximum score is 20), but still, both would get the same count of gained items by the Fenn and Hambrick's procedure.

[iii] And the fact that, in both papers (Fenn & Hambrick, 2013, 2015), the sleep and wake groups were (on average) matched for number of items recalled at test does not solve the problem. What counts is how many items can theoretically be maintained vs. gained at the participant level. Imagine a wake group of two subjects, with respectively 4 and 6 out of 10 as number of items recalled at test, and a sleep group of also two, with scores of 3 and 7. Although their average test scores are identical, a subsequent gain of three items for every participant actually mean an average proportion gain of .63 for the wake group, but one of .71 for the sleep group. Clearly, unless the groups are matched pairwise, raw item counts would miss out on an effect of sleep.

[iv] The same analysis on raw item counts detected an effect of sleep on gained items (sleep: 3.1 vs. wake: 1.3; $F(1,68) = 23.33$, $p < .0001$), but missed out on the corresponding effect on maintained items (sleep: 3.9 vs. wake: 2.8; $F(1,68) = 1.68$, $p > .19$; Group x Item trajectory: $F(1,68) = 7.17$, $p < .01$).

[v] The same analysis on raw item counts showed a similar pattern to normalized proportions, with no effect of sleep on maintained items (sleep: 9.33 vs. wake: 9.60; $F < 1$), but one on gained items (sleep: 2.13 vs. wake: 1.1; $F(1,58) = 9.22$, $p < .004$; Group x Item trajectory: $F(1,68) = 7.52$, $p < .009$).

[vi] Note that the latter analysis should be taken with caution, as it required inclusion of all participants tested by TPSWG.