The Sleep Benefit in Episodic Memory: Investigating Underlying Mechanisms

Sabrina Berres

Inaugural Dissertation
Submitted in partial fulfillment of the requirements for the degree Doctor of Social Sciences in the Graduate School of Economic and Social Sciences (GESS) at the University of Mannheim.

Supervisor:
Prof. Dr. Edgar Erdfelder

Dean of the School of Social Sciences:
Prof. Dr. Michael Diehl

Academic Director of the CDSS:
Prof. Thomas Gschwend, Ph.D.

Thesis Reviewers:
Prof. Dr. Arndt Bröder
Prof. Dr. Thorsten Meiser

Defense Committee:
Prof. Dr. Arndt Bröder
Prof. Dr. Edgar Erdfelder
Prof. Dr. Beatrice G. Kuhlmann
Prof. Dr. Thorsten Meiser

Thesis Defense:
June 19, 2023

Contents

Abstract

Sleep not only benefits health and well-being but is also considered a prerequisite for proper episodic memory functioning. Even though almost a century of research has documented a positive effect of sleep in episodic memory, our knowledge about the underlying mechanisms is still limited. The overall goal of the present thesis is to improve our understanding of the mechanisms underlying the sleep benefit by introducing an integrative framework that incorporates prominent theories from neuroscience and psychology.

Because a quantification of the sleep benefit and a systematic investigation of potential moderator variables has been missing, I addressed these points in a comprehensive, up-to-date meta-analysis. The results revealed a robust effect of moderate size. Furthermore, sleep benefits in episodic memory are significantly moderated by retrieval procedures, the definition of the dependent variable, and sleep study designs. On the basis of the meta-analytic results, I developed an integrative framework that provides a broader and more comprehensive explanation of the sleep benefit than earlier theories. Essentially, it combines active systems consolidation, opportunistic consolidation, and interference accounts (see Article 1; Berres & Erdfelder, 2021).

Using multinomial processing tree (MPT) models, the two additional original research articles of this thesis focus on testing core assumptions of the theories incorporated in the integrative framework. This methodological approach allows a more thorough investigation of hypothesized underlying mechanisms by providing uncontaminated measures and disentangling the mechanisms' joint contribution to the sleep benefit.

According to the active systems consolidation account, information salience induced by encoding strength and test expectation should moderate sleep benefits in episodic memory storage. However, previous research of Howe (1970) and McGeoch (1929) suggests that encoding strength also affects interference-based sleep benefits in memory retrieval. In line with the integrative framework, the results showed that sleep improves both episodic memory storage and retrieval. Moreover, whereas encoding strength moderates sleep benefits in retrieval rather than in storage, test expectation appears to have no effect on neither storage nor retrieval (see Article 2; Berres & Erdfelder, 2023).

The active systems consolidation account also predicts that sleep improves source memory for retention intervals of up to 12 hr (Inostroza & Born, 2013; Klinzing et al., 2019). However, previous empirical evidence for sleep benefits in source memory is quite mixed. One of the possible reasons for the conflicting results is that the source memory measures used likely confound item memory, source memory, and guessing. Therefore, I reexamined the sleep-strengthens-source-memory hypothesis with a multinomial modeling approach. Whereas findings for sleep benefits in item memory were mixed, results for source memory were in line with the active systems consolidation account. Specifically, sleep within a 12-hr retention interval improved source memory for spatial positions (see Article 3; Berres et al., 2023).

Hence, the three articles included in this thesis indicate that both active systems consolidation and interference accounts are required to fully explain the pattern of results. Moreover, the meta-analytical results provide guidance to increase the replicability of sleep-induced episodic memory benefits. Finally, I also demonstrated the power of multinomial modeling in investigating underlying mechanisms of the sleep benefit. In sum, this thesis improves our understanding of the underlying mechanisms and contributes to the theoretical progress in sleep and memory research.

Articles

This cumulative thesis is based on one published article and two articles submitted for publication. The articles will be discussed in the chronological order of development. In the main text, I provide an overview of the articles, relate them to each other, and discuss them in terms of the integrative framework. Note that details regarding theories, methods, and results are not reiterated here and can be found in the articles appended to this thesis.

ARTICLE 1

Berres, S., & Erdfelder, E. (2021). The sleep benefit in episodic memory: An integrative review and a meta-analysis. *Psychological Bulletin, 147*(12), 1309–1353. https://doi.org/10.1037/bul0000350.

ARTICLE 2

Berres, S., & Erdfelder, E. (2023). *Is it all about storage? Effects of encoding strength and test expectation on the sleep benefit in episodic memory* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

ARTICLE 3

Berres, S., Erdfelder, E., & Kuhlmann, B. G. (2023). *Does sleep benefit source memory? Investigating 12-hr retention intervals with a multinomial modeling approach* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

During my dissertation, I have also worked on another article investigating underlying mechanisms of the sleep benefit in episodic memory. As I am not the main contributor to this article, it is not included in the present thesis. Nevertheless, I will refer to this article in the main text, because it introduces the multinomial processing tree model applied in Article 2 and contains relevant empirical evidence. Please refer to Appendix A of Article 2 (Berres & Erdfelder, 2023) for a detailed description of the model.

Erdfelder, E., Berres, S., Quevedo Pütter, J., & Küpper-Tetzel, C. E. (2022). *Why does sleep improve episodic memory? An encoding-maintenance-retrieval analysis* [Manuscript under revision]. Department of Psychology, University of Mannheim.

“There is nothing so practical as a good theory.”
– *Kurt Lewin* (1952, p. 169)

1 Introduction

In its long history, research on the sleep benefit in episodic memory has produced not only a considerable amount of literature and empirical findings but also a variety of theories about the underlying mechanisms that are still vigorously debated (e.g., Antony & Schapiro, 2019; Ellenbogen et al., 2006; Yaroush et al., 1971; Yonelinas et al., 2019). As such, episodic memory is the memory for past events, experiences, or information and their respective context, such as location and time (Tulving, 2002). It comprises the ability to encode, store, and retrieve episodic information.

The surge of sleep and memory research started with Ebbinghaus (1885) who observed better memory performance for approximately 9 hr and for 1 day after learning than predicted by his logarithmic forgetting function. He interpreted this finding as an unsystematic error in his observations (see pp. 85–109). Since Ebbinghaus' interpretation, the explanation for the irregularity in forgetting rates he observed evolved into detailed theories about sleep-induced memory benefits, which even motivate memory-enhancing sleep intervention techniques today (Feld & Diekelmann, 2020). Most of these theories can be assigned to one of the following four theoretical accounts:

Active Systems Consolidation Account

According to the active systems consolidation account, recently encoded and therefore labile memories are converted into more stable long-term memory representations during sleep, thereby improving episodic memory storage (e.g., Diekelmann & Born, 2010; Dudai, 2004, 2012; Dudai et al., 2015; Klinzing et al., 2019; Rasch & Born, 2013). There are various theories that explain sleep benefits in episodic memory by memory consolidation. One of these theories is the sequential hypothesis which predicts that memories are consolidated in two consecutive steps during slow-wave sleep (SWS) and rapid eye movement (REM) sleep (Ambrosini & Giuditta, 2001; Giuditta, 2014; Giuditta et al., 1995). Another theory is the synaptic homeostasis hypothesis according to which sleep restores cellular homeostasis by renormalizing synaptic connections that were strengthened during wakefulness (Cirelli & Tononi, 2015; Tononi & Cirelli, 2003, 2006, 2014, 2020).

The present thesis, however, focuses on the active systems consolidation hypothesis, because it incorporates aspects of various consolidation theories—including the sequential and synaptic homeostasis hypothesis. As such, this theory predicts that during wakefulness components of a memory representation are formed, distributed across various neocortical brain areas, and bound together into a unique episodic memory representation by the hippocampus. During subsequent sleep, especially during SWS, the hippocampal memory representation is replayed by the reactivation of specific neuronal firing patterns (Klinzing et al., 2019; Lewis & Durrant, 2011; O'Neill et al., 2010; Pfeiffer, 2020; Wilson & McNaughton, 1994). In parallel, the separate components of the memory representation are also reactivated in the various neocortical brain areas. These local synaptic upscaling processes which strengthen synaptic connections and stabilize memory representations are embedded in global synaptic downscaling which renormalizes the strength of synaptic connections across all cortical and subcortical areas by diminishing neuronal firing rates (Feld & Born, 2017; Klinzing et al., 2019).

Finally, the combination of local synaptic upscaling and global synaptic downscaling in the hippocampus and neocortex should eventually result in a net strengthening of episodic context-bound hippocampal memory representations for retention intervals of up to 12 hr and more gist-like decontextualized neocortical memory representations for longer retention intervals (e.g., 3 days; Klinzing et al., 2019). Taken together, the consolidation account in general, and the active systems consolidation hypothesis in particular, propose an active role of sleep such that sleep strengthens episodic memory storage for 12-hr retention intervals (Born & Wilhelm, 2012; Diekelmann & Born, 2010; Feld & Born, 2017; Inostroza & Born, 2013; Klinzing et al., 2019; Rasch & Born, 2013).

Opportunistic Consolidation Account

In contrast to the active systems consolidation account, the opportunistic consolidation account predicts that memory consolidation is not sleep specific. Rather, any other state that provides a beneficial environment for memory consolidation by reducing retroactive interference (e.g., quiet wake, alcohol, benzodiazepines) should improve episodic memory storage (e.g., Ellenbogen et al., 2006; Mednick et al., 2011; Wixted, 2004, 2005). Specifically, it is assumed that memory consolidation can be disturbed by new learning, because both consolidation and learning processes rely on the same limited hippocampal resources (Wixted, 2004, 2005). Hippocampal memory consolidation should occur whenever the opportunity is provided, that is, whenever interference from subsequent learning (retroactive interference) is reduced. During sleep, for example, new learning is virtually absent. As a consequence, more hippocampal resources are available for memory consolidation of information learned before sleep (Wixted, 2004, 2005). Hence, the opportunistic consolidation account proposes a passive role of sleep in episodic memory storage.

Interference Account

Like the opportunistic consolidation account, the interference account proposes a passive role of sleep in episodic memory. However, there is a crucial difference between both theoretical accounts: Whereas the opportunistic consolidation account predicts that sleep improves episodic memory storage, the interference account predicts that sleep improves episodic memory retrieval. Specifically, according to the interference account, sleep compared to wakefulness protects memories from retroactive interference, thereby improving episodic memory retrieval (e.g., Ecker, Brown, & Lewandowsky, 2015; Ecker, Tay, & Brown, 2015; Ekstrand, 1967; Jenkins & Dallenbach, 1924; Yonelinas et al., 2019). Because new learning is virtually absent during sleep, retroactive interference is reduced compared to wakefulness. Importantly, previous research has shown that retroactive interference is primarily caused by retrieval and not by storage problems (see, for example, Tulving & Psotka, 1971). It is therefore assumed that the reduction of retroactive interference during sleep supports episodic memory retrieval. In sum, the interference account proposes a passive role of sleep in episodic memory retrieval by protecting memories from retroactive interference.

Circadian-Effects Account

The circadian-effects account predicts that observed sleep benefits in episodic memory are due to circadian effects on encoding and retrieval processes, because sleep and wakefulness are naturally confounded with diurnal times of learning and testing (e.g., Abel & Bäuml, 2013a; Bäuml et al., 2014; Erdfelder et al., 2022; Fenn & Hambrick, 2012, 2013; Payne et al., 2008). To elaborate, a huge bulk of research on sleep and memory compares nighttime sleep with daytime wakefulness. This sleep study design confounds sleep and wakefulness with diurnal times of learning and testing (e.g., Abel & Bäuml, 2013a; Bäuml et al., 2014; Erdfelder et al., 2022; Fenn & Hambrick, 2012, 2013; Payne et al., 2008).

To check for circadian effects, researchers have therefore introduced morning and evening control conditions in which encoding and retrieval follow in close succession parallel to the sleep and wake conditions' respective learning and testing time. Consider, for example, that a specific daytime (e.g., evening) is beneficial for one process (e.g., encoding) but detrimental for the other (e.g., retrieval), and for another daytime (e.g., morning) it is the other way around. This would lead to a null overall effect of daytime in the control conditions, a beneficial daytime effect in the sleep condition, and a detrimental effect of daytime in the wake condition. Thus, the typical result pattern found in sleep and wake conditions of natural sleep studies can also be explained by phase-shifted circadian rhythms without assuming a causal effect of sleep on memory.

Although significant progress and important insights into the neurocognitive mechanisms underlying the sleep benefit have been made in past decades, sleep and memory research is still marked by scientific controversies and theories that require further refinement and more rigorous testing. One way to investigate the underlying mechanisms of the sleep benefit in episodic memory more thoroughly is the application of appropriately designed multinomial processing tree (MPT) models (for reviews, see Batchelder & Riefer, 1999; Erdfelder et al., 2009).

1.1 Investigating Underlying Mechanisms

Following this approach, Erdfelder et al. (2022) adopted the Encoding-Maintenance-Retrieval (EMR) multinomial model (Küpper-Tetzel & Erdfelder, 2012) to a typical sleep study design for paired-associate learning (for details see Appendix A of Article 2; Berres & Erdfelder, 2023). In this study design, participants learn weakly related word pairs before performing an immediate cued recall task. After a retention interval, participants perform a delayed free recall task followed by final cued recall. The authors applied the EMR model in two experiments, investigating how sleep compared to wakefulness affects episodic memory storage and retrieval.

As outlined above, sleep benefits in episodic memory storage are predicted by active systems and opportunistic consolidation accounts. To evaluate active versus passive sleep effects on memory storage, neuroimaging techniques, polysomnography, and the investigation of various states that reduce retroactive interference (e.g., quiet wake, alcohol) are required. This is, however, outside the scope of the work by Erdfelder et al. (2022) and the present thesis as both compare sleep versus wakefulness by using behavioral data. Although an additional passive role of sleep in memory storage as proposed by the opportunistic consolidation account cannot be ruled out completely, previous research has clearly shown that sleep actively improves episodic memory storage (for a review of neurocognitive evidence, see Article 1; Berres & Erdfelder, 2021). For this reason, Erdfelder et al. (2022) and the present thesis focus on the active systems consolidation account in terms of sleep benefits in episodic memory storage.

In Experiment 1, Erdfelder et al. (2022) evaluated the active systems consolidation account, the interference account, and the circadian-effects account. According to the active systems consolidation account, memory storage (parameter m) should be larger when participants sleep during a 12-hr retention interval than when they stay awake. By contrast, the interference account predicts a sleep benefit in memory retrieval (parameter r_f). Finally, the circadian-effects account predicts phase-shifted circadian rhythms for encoding and retrieval success, that is, better encoding (parameter e) in the evening than in the morning, which should be reversed for retrieval (parameter r_f).

To test these hypotheses, a lab experiment ($N = 40$) with a randomized 2×2 between-subjects design, including study time (9 a.m. vs. 9 p.m.) and length of the retention interval (6 min vs. 12 hr) as independent variables was conducted. The authors found that the length of the retention interval affects storage (m) and free recall retrieval (r_f), but not encoding parameters (e). Taken together, the results of Experiment 1 suggest that sleep compared to wakefulness benefits both storage and retrieval, thereby providing preliminary support for the active systems consolidation and interference account. By contrast, there was no empirical evidence for circadian effects which is in line with previous research (e.g., Abel & Bäuml, 2012, 2013a, 2013b; Bäuml et al., 2014; Fenn & Hambrick, 2013).

In Experiment 2, the authors aimed not only to replicate the results for the 12-hr retention interval of Experiment 1 with a larger sample but also to provide additional evidence for the validity of the EMR model. Specifically, the model's capacity to disentangle storage and retrieval processes was tested by providing half of the participants with a list of semantic category labels as retrieval cues for the word pairs in the delayed free recall task. This manipulation should selectively affect the retrieval of stored associations in free recall (r_f) by increasing retrieval success in the condition with present category labels, leaving encoding (e) and storage processes (m) unaffected.

A lab experiment ($N = 60$) with a randomized 2×2 between-subjects design, including study time (9 a.m. vs. 9 p.m.) and category labels (absent vs. present) as independent variables was conducted. Replicating the results of the 12-hr conditions in Experiment 1, the model-based analysis showed statistically significant better memory storage (m) and retrieval (r_f) after sleep than wakefulness, supporting the active systems consolidation and the interference account, respectively. Furthermore, introducing category labels as retrieval cues in free recall increased memory retrieval (r_f) but left encoding (e) and storage (m) unaffected, thereby providing additional evidence for the EMR model's psychological validity.

To summarize, both experiments show that the EMR model is an appropriate tool to investigate the sleep benefit in episodic memory. In addition the results of the two experiments suggest that sleep compared to wakefulness not only actively strengthens episodic memory storage but also passively supports episodic memory retrieval. Importantly, the findings of Experiment 1 are clearly at odds with the circadian-effects account, ruling out this alternative explanation. Moreover, previous research showed compelling neurocognitive evidence for an active role of sleep in episodic memory storage (for a review, see Article 1; Berres & Erdfelder, 2021). For these reasons, the present thesis focuses on active systems consolidation and interference accounts.

1.2 An Integrative Framework

So far, theoretical debates on the mechanisms underlying the sleep benefit have focused on either-or questions (e.g., “Does sleep improve episodic memory by boosting memory consolidation or by reducing retroactive interference?”), treating theoretical explanations as mutually exclusive. By contrast, I promote an integrative, multicausal framework in my thesis. This framework was inspired by the findings of Erdfelder et al. (2022) and developed based on meta-analytic results (see Article 1; Berres & Erdfelder, 2021). Specifically, it integrates three prominent classes of theories from neuroscience and psychology, namely, the active systems consolidation, opportunistic consolidation, and interference account.

In a nutshell, the integrative framework posits that sleep compared to wakefulness after learning affects episodic memory in two ways: First, sleep actively strengthens memory storage due to memory consolidation (active systems consolidation account), perhaps additionally supported by passive protection from interference (opportunistic consolidation account). Second, sleep passively improves memory retrieval by reducing retroactive interference (interference account). As such, the assumption that sleep compared to wakefulness fosters both episodic memory storage and retrieval in parallel offers a more comprehensive explanation of the sleep benefit than previous more fine-graded theories do. By bridging the gap between neuroscientific and psychological accounts, the proposed integrative framework is a promising basis to derive additional predictions, guide future research, and stimulate methodological innovations (Berres & Erdfelder, 2021).

In the following, I briefly summarize the first article which provides a review of hypothesized underlying processes, a comprehensive quantification of the sleep benefit, a systematic investigation of potential moderator variables, and the integrative framework described above, which is based on the meta-analytic results. I then provide an overview of each of the two subsequent articles in which core assumptions of the theories incorporated in the integrative framework are tested with a multinomial modeling approach. Specifically, the second article tests the prediction of the active systems consolidation account that information salience manipulated via encoding strength and test expectation moderates the sleep benefit in episodic memory storage. It also addresses the question whether effects of encoding strength additionally moderate sleep benefits in episodic memory retrieval (interference account). The third article aims to generalize previous research on the sleep benefit in item-item associations to item-context associations. According to the active systems consolidation account, sleep should improve source memory for retention intervals of up to 12 hr (Inostroza & Born, 2013; Klinzing et al., 2019). Because previous findings are quite mixed, I reexamined the sleep-strengthens-source-memory hypothesis with a multinomial modeling approach. Finally, I will discuss limitations of the present thesis and directions for future research on sleep and episodic memory.

In sum, I hope that this thesis contributes to the development of a "good theory" in Lewin's (1952) sense. That is, a theory that helps us understand not only why sleep boosts memory but also how to maximize the positive effects of sleep. Indeed, such a theory is of considerable value for developing effective memory-enhancing sleep interventions in dementia treatment or psychotherapy (e.g., Blackman et al., 2021; Feld & Diekelmann, 2020; Goerke et al., 2017; Pace-Schott et al., 2015a, 2015b).

2 The Sleep Benefit in Episodic Memory

Berres, S., & Erdfelder, E. (2021). The sleep benefit in episodic memory: An integrative review and a meta-analysis. *Psychological Bulletin, 147*(12), 1309–1353. https://doi.org/10.1037/bul0000350.

Although Ebbinghaus (1885) was the first to document the positive effect of sleep on memory, it took about 40 years for the observed memory advantage to be attributed to sleep. Specifically, in their pioneering work, Jenkins and Dallenbach (1924) investigated memory for nonsense syllables of two participants in a free recall task by varying the time of learning (i.e., day or night) and the length of the retention interval (i.e., 1, 2, 4, or 8 hr). Comparing retention intervals of equal length filled with either sleep or wakefulness, the authors found clear evidence for better memory performance after sleep. Today, the sleep benefit in episodic memory is considered a well-established effect. However, recent findings and failed replication attempts have cast doubt on its robustness and theoretical foundation (for a review, see Cordi & Rasch, 2021).

2.1 An Integrative Review and a Meta-Analysis

Critically, despite the wealth of research on the sleep benefit in episodic memory, an integrative review of hypothesized underlying mechanisms, a comprehensive quantification of the sleep benefit, and a systematic investigation of potential moderator variables had been missing. To fill this gap, these three points were addressed as follows:

First, because there is still a lot of controversy about the underlying neurocognitive mechanisms, an exhaustive review of prominent explanations of the sleep benefit originating from neuroscience and psychology was provided. Specifically, after outlining active systems consolidation, opportunistic consolidation, and interference accounts, empirical evidence from neuroimaging and polysomnography research was reviewed to evaluate the active systems consolidation and opportunistic consolidation account. This review showed compelling evidence for the active systems consolidation account and highlighted that more research on other states of reduced retroactive interference (e.g., quiet wake, alcohol, benzodiazepines) is needed to evaluate the opportunistic consolidation account. Next, hypotheses for potential moderator variables were derived from active systems consolidation and interference accounts to test the predictions of both theories in a meta-analysis on behavioral measures of the sleep benefit. As such, the review and the meta-analysis laid the foundation for the integrative framework proposed in this article.

Second, as the typical size of the sleep benefit in episodic memory across studies was still unknown, an extensive up-to-date meta-analysis on healthy drug-free human individuals was conducted to quantify its size. Specifically, the meta-analysis is based on 823 effect sizes from 271 independent samples that were reported in 177 articles published between 1967 and 2019. Using an intercept-only three-level model with robust variance estimation (RVE) to account for dependencies among effect sizes due to multiple outcome reporting and multiple group comparisons in primary studies, I found a moderate sleep benefit in episodic memory (Hedges' $g = 0.44$).

However, this effect might be overestimated due to selective reporting in the primary literature, leading to an overrepresentation of significant results in the hypothesized direction. Importantly, although the overall effect of sleep on episodic memory was reduced, it was still significant when accounting for selective reporting by using a modified variant of Egger's linear regression (Hedges' $g = 0.28$; Pustejovsky & Rodgers, 2019; Rodgers & Pustejovsky, 2020) and by applying the three-parameter selection model (3PSM; Vevea & Hedges, 1995) to 20 subsets of independent effect sizes ($Mdn_g = 0.41$).

As such, the sleep benefit in episodic memory appears to be quite similar in size to sleep benefits in prospective (Cohen's $d = 0.41$; Leong et al., 2019) and motor memory (Cohen's $d = 0.44$, Pan & Rickard, 2015; Hedges' $g = 0.43$, Schmid et al., 2020) and comparable to typical well-established memory effects (e.g., testing effect: Hedges' $g = 0.50$, Rowland, 2014; retrieval-induced forgetting effect: Hedges' $g = 0.35$, Murayama et al., 2014). Consequently, the sleep benefit in episodic memory can be considered a robust effect, comparable in size to other well-established effects in human memory research.

By quantifying the sleep benefit, this work helps researchers to plan studies that are sufficiently powered, thereby increasing replicability in the future. Furthermore, it also informs about the practical significance of the sleep benefit in episodic memory, for example, for aging and mental health research (e.g., Feld & Diekelmann, 2020; Goerke et al., 2017; Gui et al., 2017; Lo et al., 2016; Mander et al., 2017; Pace-Schott et al., 2015a, 2015b; Palagini et al., 2013; Scullin, 2017; Scullin & Bliwise, 2015; Wilhelm et al., 2012).

Third, employing multilevel meta-regressions with robust variance estimates, potential moderator variables on the sleep benefit were examined systematically. To investigate beneficial sleep effects on episodic memory, researchers have used a wide variety of populations (e.g., children, older adults), materials (e.g., word pairs, pictures), sleep study designs (e.g., naps, total sleep deprivation), and dependent measures (e.g., cued recall, recognition). However, the impact of these potential moderators had not yet been studied systematically and in detail. It was therefore crucial to examine potential moderator effects with up-to-date meta-analytical methods as aggregating empirical findings across various studies allows unique insights that can hardly be obtained from single experiments.

In total, 16 potential moderators from five topic areas were investigated: (a) population characteristics (age), (b) aspects of the encoding situation (emotionality, restudy, testing, feedback, encoding instruction, directed forgetting instruction), (c) aspects of memory assessment (memory test, definition of the dependent variable), (d) aspects of the study design (stimulus material, sleep study design, length of the retention interval, experimental design), and (e) sleep parameters (total sleep time [TST], slow-wave sleep [SWS], rapid eye movement [REM] sleep). Furthermore, two control variables were included in the meta-analysis to investigate whether randomization in primary studies and the way effect sizes were extracted from primary studies affect the results. In addition, a series of sensitivity analyses were performed which showed that the results of the meta-analysis are by and large robust.

2.2 Developing an Integrative Framework

Although the meta-analysis was limited to variables for which sufficient data was available for joint moderator analyses, a wide range of potential moderators was examined, revealing four important findings: First, studying the stimulus material multiple times instead of just once appears to increase sleep-induced memory benefits, as predicted by the interference account.

Second, the largest sleep benefit emerged in free recall, followed by cued recall, and lastly recognition tasks. Whereas almost all performed analyses showed this result pattern descriptively, it was statistically significant for word materials (i.e., single words and word pairs) only. Thus, at least for word materials, sleep not only boosts episodic memory as predicted by the active systems consolidation account, but also reduces retroactive interference in line with the interference account, leading to larger sleep benefits in memory tasks that rely more on memory retrieval (cf. Dyne et al., 1990; McKinney, 1935; Postman, 1952).

Third, sleep benefits in episodic memory appear to be larger for differences between immediate and delayed test performance than for delayed test performance, suggesting that researchers should favor pre-post difference measures that control for individual differences by design to increase the power of statistical tests. Fourth, sleep study designs without constraints on SWS (e.g., natural sleep, nighttime naps) yielded larger sleep benefits than sleep study designs with SWS-constraints (e.g., SWS-deprived sleep, daytime naps).[1] This pattern fits well to the assumption of the active systems consolidation account that SWS is important for memory consolidation while it also urges caution in generalizing findings across different sleep study designs.

Overall, the results provide initial support for both active systems consolidation and interference accounts, suggesting that the two theories should not be considered as mutually exclusive. Rather, it appears that sleep fosters memory storage by memory consolidation and improves memory retrieval by reducing retroactive interference simultaneously which is at the heart of the integrative framework developed based on the meta-analytic results. Besides evaluating and integrating major theoretical accounts of the sleep benefit (i.e., active systems consolidation, opportunistic consolidation, and interference accounts) to better understand the underlying processes, this work also helps researchers design informative replication studies and explain unexpected results.

2.3 Implications for Future Research

In sum, this article presents the first comprehensive meta-analysis on the sleep benefit in episodic memory. Compared to other meta-analyses on sleep and memory (e.g., Hu et al., 2020; Schäfer et al., 2020; Schmid et al., 2020), it contains the largest number of studies, effect sizes, and moderator variables so far. The present meta-analysis also provides important theoretical insights on the sleep benefit and offers practical guidance to increase its replicability in future research. However, the integrative framework and its incorporated theories require more rigorous testing. Specifically, two important core assumptions of the active systems consolidation account were not sufficiently addressed in the meta-analysis and require further evaluation:

First, information salience—manipulated via encoding strength and test expectation—should moderate sleep benefits in episodic memory storage. However, previous research of Howe (1970) and McGeoch (1929) suggest that encoding strength also affects interference-based sleep benefits in memory retrieval. Second, sleep should improve source memory for retention intervals of up to 12 hr (Inostroza & Born, 2013; Klinzing et al., 2019). However, previous empirical evidence for sleep

---

[1] In a sleep study design with natural sleep, participants in the sleep condition learn the material in the evening and are tested after a night of sleep in the morning, whereas in the wake condition the time of learning and testing is reversed. In nighttime and daytime nap designs, participants in the sleep condition nap during the night (i.e., after 7 p.m.) and during the day (i.e., before 7 p.m.), respectively, whereas participants in the wake condition stay awake throughout this time. In a sleep study design with SWS deprivation, participants either sleep in the second part of the night to reduce SWS or are awakened every time the first epoch of SWS occurs throughout a full night of sleep. This sleep condition is compared to a wake condition with an equally long period of daytime wakefulness.

benefits in source memory is mixed, in part perhaps due to source memory measures that confound item memory, source memory, and guessing.

But how can sleep benefits in episodic memory storage and retrieval or in source memory be disentangled and appropriately measured? As such, multinomial processing tree models provide an elegant solution to this problem. In the next section, I first introduce this model class before turning to the two articles in which the multinomial modeling approach was used to test the core assumptions outlined above.

3 A Multinomial Modeling Approach

Multinomial processing tree (MPT) models are substantively motivated stochastic models that explain observed categorical behavioral data by estimating probabilities for the occurrence of hypothesized latent processes (Batchelder & Riefer, 1999; Erdfelder et al., 2009; Riefer & Batchelder, 1988). Specifically, these models assume that observed responses originate from a finite set of discrete processing states that can be explicitly specified by branches in a tree structure. Thus, each branch of a processing tree represents a specific hypothesized sequence of latent processing states that results in a particular response, whereas a particular response can arise from one or more branches. By assuming that observed response frequencies follow multinomial distributions modeled by a processing tree, the contribution of each hypothesized latent cognitive process can be estimated and tested (Batchelder & Riefer, 1999; Erdfelder et al., 2009; Hu & Batchelder, 1994; Riefer & Batchelder, 1988).

Responses in an episodic memory task are rarely assumed to result from a single cognitive process. Consider for example a standard source-monitoring task in which participants study items from two sources and subsequently indicate whether items were presented previously, and if so, in which source (e.g., Bayen et al., 1996; Kuhlmann et al., 2021; Murnane & Bayen, 1996). The observed responses may result from multiple cognitive processes, such as memory and guessing processes, rather than a single process. In contrast to memory performance measures (e.g., hit rates) which likely confound memory and guessing processes (e.g., Bröder & Meiser, 2007; Murnane & Bayen, 1996), MPT models provide an elegant solution to decompose the contributions of assumed latent cognitive processes. As formal mathematical measurement models that are well developed and understood, MPT models are powerful tools to rigorously test theoretical predictions (Kuhlmann et al., 2019). In fact, once the validity of a MPT model is established, its parameter estimates can be interpreted as uncontaminated measures of the hypothesized latent cognitive processes (Hütter & Klauer, 2016). Due to these advantages, MPT models have gained popularity and provided valuable insights into the cognitive processes underlying human behavior (for reviews, see Batchelder & Riefer, 1999; Erdfelder et al., 2009; Hütter & Klauer, 2016).

Model Structure

To illustrate the structure of MPT models, consider the two-high-threshold multinomial model of source monitoring (2HTSM) in Figure 1. This model was applied in Experiment 1 of Article 3 to disentangle cognitive processes that are likely confounded in commonly used source memory measures. In a standard source-monitoring task, participants studied items (i.e., pictures) that were presented on either the left or right side of a computer screen. Subsequently, the presented items (i.e., targets) were mixed with new items (i.e., distractors) and participants were asked to indicate whether the item was presented previously, and if so, in which source (e.g., Bayen et al., 1996; Murnane & Bayen, 1996).

As discussed above, observed responses are likely due to multiple latent cognitive processes rather than one. If participants recognize a target item and identify its source correctly, they will provide the correct answer. However, they will also answer correctly by recognizing the target item and guessing the correct source in which the target was presented. By explicitly specifying the hypothesized latent cognitive processes underlying the observed responses, the 2HTSM provides uncontaminated measures for item memory, source memory, and guessing. The probability that each branch within a tree will result in a particular response is calculated by taking the product of the (conditional) probabilities for all processing states along the branch. Because multiple branches can lead to the same response, the corresponding response probability is the sum of all branches that end in this response category (Batchelder & Riefer, 1990).

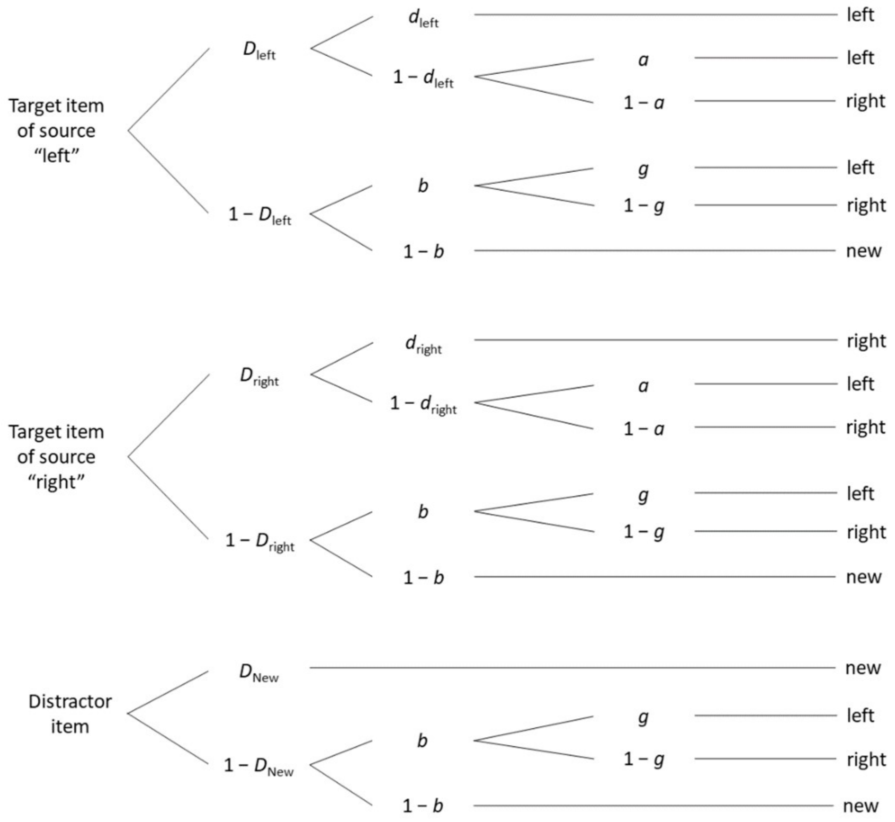
Model Fit and Model Validity

MPT models are most often fitted to empirical data by complete or partial pooling methods implemented in appropriate software (see, for example, Heck et al., 2018; Moshagen, 2010). In the complete pooling approach, maximum likelihood (ML) parameter estimation is used to analyze observed category frequencies aggregated across participants and items (Hu & Batchelder, 1994). In the partial pooling method, a Bayesian approach that relies on Markov-chain Monte Carlo (MCMC) methods is employed to estimate individual and group-level MPT-parameters. By combining information on the individual and group level, this approach explicitly accounts for potential individual differences (Heck et al., 2018). In this thesis, both methods are used to check the robustness of the results with respect to different distributional assumptions involved in complete and partial pooling. Importantly, all results reported in the two articles below appear to be robust.

Using a goodness-of-fit-test, the fit between empirical data and the MPT model can be assessed. Furthermore, the validity of the MPT model's assumptions can be established by showing that experimental manipulations selectively affect specific parameters (Erdfelder et al., 2009; Hütter & Klauer, 2016) or by demonstrating functional dissociation with continuous covariates (Bott et al., 2020). Once, the MPT model's identifiability (i.e., a one-to-one mapping of parameter values to observed data), fit, and validity is established, parameter estimates can be interpreted as uncontaminated measures, and hypothesis tests can be performed (Erdfelder et al., 2009; Hütter & Klauer, 2016). In the two articles discussed below, appropriately designed MPT models were applied to test core assumptions of theories incorporated in the integrative framework.

Figure 1

Two-High-Threshold Multinomial Model of Source Monitoring (2HTSM) Adapted to the Spatial Position Source Manipulation Used in Experiment 1 of Article 3



Note. D_{left} = probability of correctly identifying a target item in source "left" as "old"; D_{right} = probability of correctly identifying a target item in source "right" as "old"; D_{New} = probability of correctly identifying a distractor item as "new"; d_{left} = probability of correctly identifying the source of a target item in source "left"; d_{right} = probability of correctly identifying the source of a target item in source "right"; a = probability of guessing that a correctly identified target item is from source "left"; b = probability of guessing that an item is "old"; g = probability of guessing that an unrecognized item is from source "left" if it was guessed to be "old". Following Submodel 5a of the 2HTSM (Bayen et al., 1996), two equality constraints—$D_{left} = D_{right} = D_{New}$ and $d_{left} = d_{right}$—were imposed on the model parameters. Adapted from "Source Discrimination, Item Detection, and Multinomial Models of Source Monitoring", by U. J. Bayen, K. Murnane, and E. Erdfelder, 1996, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(1), p. 202 (https://doi.org/10.1037/0278-7393.22.1.197). Copyright 1996 by the American Psychological Association.

3.1 Investigating Effects of Encoding Strength and Test Expectation

Berres, S., & Erdfelder, E. (2023). *Is it all about storage? Effects of encoding strength and test expectation on the sleep benefit in episodic memory* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

One core prediction of the active systems consolidation account is that specific aspects of the encoding situation, such as encoding strength or test expectation, selectively enhance memory consolidation during sleep and thus episodic memory storage. Specifically, it is assumed that memories are tagged during or shortly after encoding (e.g., Barco et al., 2008; Frey & Morris, 1998; Morris, 2006; Redondo & Morris, 2011). These tags should boost memory consolidation during sleep, resulting in better memory storage (e.g., Antony & Paller, 2017; Bennion et al., 2016; Stickgold & Walker, 2013).

However, recent evidence suggests that sleep not only improves episodic memory storage through memory consolidation but also episodic memory retrieval by reducing retroactive interference (Berres & Erdfelder, 2021; Erdfelder et al., 2022). In addition, previous research showed that retroactive interference tends to decrease with higher study efforts (Howe, 1970; McGeoch, 1929). Because sleep benefits in episodic memory retrieval rely on relatively strong retroactive interference during wakefulness, the research by Howe (1970) and even earlier research by McGeoch (1929) suggests that sleep-induced retrieval benefits decrease with increasing encoding strength. Hence, there might be an additional effect of encoding strength on sleep benefits in memory retrieval.

An Encoding-Maintenance-Retrieval Analysis

To evaluate the active systems consolidation account, information salience was manipulated during (encoding strength, Experiment 1) and after learning (test expectation, Experiment 2). Moreover, the Encoding-Maintenance-Retrieval (EMR) model (Erdfelder et al., 2022; Küpper-Tetzel & Erdfelder, 2012) was applied to investigate whether information salience affects sleep benefits in episodic memory storage, retrieval, or both. As such, the EMR model is based on an extended free-then-cued recall paradigm: First, participants study weakly associated word pairs (e.g., “candy – bread”) and are then asked to recall the second word of a word pair in an immediate cued recall task that provides the first word as a memory aid (i.e., “candy – ??”). Please refer to Appendix A of Article 2 (Berres & Erdfelder, 2023) for a detailed description of the model.

After a retention interval (filled with either a period of sleep or wakefulness), participants perform a delayed free recall task, in which the studied stimuli have to be retrieved without cues. This is followed by a delayed cued recall task, which is similar to the immediate cued recall task (Küpper-Tetzel & Erdfelder, 2012, see also Küpper-Tetzel et al., 2014). Because the EMR model allows to estimate the probabilities of successfully maintaining encoded associations across the retention interval (parameter m), and retrieving stored associations in free recall (parameter r_f), it is a powerful tool to investigate how encoding strength and test expectation affect the sleep benefit and its underlying processes. In the following, I briefly summarize the two experiments included in this article.

Investigating Effects of Encoding Strength

Experiment 1 aimed at investigating whether encoding strength manipulated in the medium to high range (i.e., $\gtrsim$ 50% mean correct immediate recall) affects sleep benefits in episodic memory storage, sleep benefits in episodic memory retrieval, or both. In line with Stickgold's (2009) inverted U-shaped function of the sleep benefit for memory strength, a larger sleep benefit in episodic memory storage for medium- compared to strong-encoding conditions was expected. Specifically, it is assumed that very strong memories should benefit less from further strengthening through memory consolidation during sleep because they already possess sufficient memory strength by definition (Denis et al., 2021; Denis et al., 2020; Petzka et al., 2021). The hypothesis outlined above was tested in a lab experiment ($N = 27$) with a randomized 2 × 2 mixed factorial design, including encoding strength (medium vs. strong encoding) as within-subject factor and wake versus sleep as between-subjects factor. As such, encoding strength was manipulated by presenting half of the stimuli once versus three times.

Replicating the model-based results of Erdfelder et al. (2022), sleep compared to wakefulness improved both episodic memory storage (parameter m) and retrieval (parameter r_f) significantly. Furthermore, in contrast to the active systems consolidation account, which predicts larger sleep benefits for medium- compared to strong-encoded memories, sleep appears to improve storage (m) for all memories equally. What is more, the model-based results showed a statistically significant larger sleep benefit in episodic memory retrieval (r_f) for medium- compared to strong-encoded memories. Specifically, whereas a clear sleep benefit emerged in the medium-encoding condition, it was absent in strong encoding due to larger retrieval probabilities in the wake condition. This is in line with what was expected based on previous research by Howe (1970) and McGeoch (1929). Overall, the results of Experiment 1 suggest that encoding strength affects sleep benefits in episodic memory retrieval more than sleep benefits in episodic memory storage. Because previous findings were typically interpreted solely in terms of the active systems consolidation account, this result has important implications for our understanding of encoding-strength effects on sleep benefits.

Investigating Effects of Test Expectation

As sketched above, tagging of memories for preferred memory consolidation during sleep can take place not only during but also shortly after encoding (e.g., Barco et al., 2008; Frey & Morris, 1998; Morris, 2006; Redondo & Morris, 2011). In Experiment 2, the active systems consolidation account was therefore further evaluated by manipulating test expectation (i.e., whether or not participants expect a delayed memory test) after learning. Another objective of Experiment 2 was to check, whether the results of the strong-encoding condition in Experiment 1 can be conceptually replicated by using a 60%-learning criterion in the immediate cued recall task.[2]

The 60%-learning criterion was also applied by Wilhelm et al. (2011) who showed that sleep compared to wakefulness improves delayed cued recall performance in expectation, but not in no-expectation conditions. The authors explained this finding by enhanced consolidation of memories that are expected to be of future relevance (Wilhelm et al., 2011). To test this interpretation, a lab experiment ($N = 70$) with a randomized 2 × 2 between-subjects design, including test expectation (no expectation vs. expectation) and wake versus sleep as independent variables was conducted. Replicating the results of the strong-encoding condition in Experiment 1, the model-based results showed a statistically significant sleep benefit in episodic memory storage (m), but not in retrieval (r_f) for expectation and no-expectation conditions. Notably, the results showed no evidence for a larger sleep benefit in episodic memory storage (m) for stimuli expected to be tested in the future. This is clearly at odds with Wilhelm et al.'s (2011) explanation but in line with recent findings by Ashton and Cairney (2021) and Reverberi et al. (2020) suggesting that test expectation does not affect the sleep benefit in episodic memory.

To conclude, the presented behavioral results for salience-cue effects on episodic memory storage as predicted by the active systems consolidation account are mixed. Nevertheless, investigating effects of encoding strength and test expectation on the sleep benefit with the EMR model considerably advanced our understanding of these complex moderator effects. Importantly, both the active systems consolidation and the retroactive interference account appear to be necessary to explain encoding-strength effects on the sleep benefit, thereby providing further support for the integrative framework.

Note that all experiments of Article 2 used word pairs (i.e., item-item associations) as stimulus material. Indeed, only few studies have investigated the sleep benefit in

---

[2] In sleep and memory research, a 60%-learning criterion is often used to ensure sufficient memory strength (e.g., Backhaus & Junghanns, 2006; Feld et al., 2013; Fenn & Hambrick, 2012, 2013; Gais & Born, 2004; Marshall et al., 2004; Plihal & Born, 1997; Wilhelm et al., 2011).

episodic memory using item-source associations. Because these studies yielded mixed results, a multinomial modeling approach was adopted in the next article to assess the involved processes more rigorously.

3.2 Reexamining Sleep Benefits in Source Memory

Berres, S., Erdfelder, E., & Kuhlmann, B. G. (2023). *Does sleep benefit source memory? Investigating 12-hr retention intervals with a multinomial modeling approach* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

The hippocampus appears to play a crucial role in episodic memory by binding the content of memories (i.e., item memory) to its unique context (i.e., source memory) during encoding (for reviews, see Eichenbaum et al., 2007; Mitchell & Johnson, 2009). As stated by the active systems consolidation account, hippocampal memory representations are replayed by reactivating specific neuronal firing patterns during subsequent sleep, especially during SWS (Klinzing et al., 2019; Lewis & Durrant, 2011; O'Neill et al., 2010; Pfeiffer, 2020; Wilson & McNaughton, 1994). This results in a strengthening of these context-bound hippocampal memory representations for relatively short retention intervals (e.g., 12 hr; cf. Jurewicz et al., 2016; Lutz et al., 2017), whereas more gist-like decontextualized neocortical memory representations are observed for longer retention intervals (e.g., 3 days; Klinzing et al., 2019).

However, previous studies investigating sleep effects on source memory for retention intervals of up to 12 hr yielded mixed results (e.g., Köster et al., 2017; Lewis et al., 2011; Mawdsley et al., 2014; van der Helm et al., 2011; Wang & Fu, 2009; Wang et al., 2017). These studies differ in several aspects, such as sleep study designs, stimulus materials, and encoding instructions. In addition, the source memory measures used in these studies also likely contribute to the inconsistent findings. As such, judgments about the origin of a memory involve multiple cognitive processes like memory, decision making, guessing, and response biases (Johnson et al., 1993). Critically, several frequently used measures of item and source memory confound these processes (e.g., Bayen et al., 1996; Bröder & Meiser, 2007; Murnane & Bayen, 1996). In this article, the core prediction of the active systems consolidation account that sleep compared to wakefulness benefits source memory for 12-hr retention intervals was therefore reexamined in two experiments using MPT models to disentangle item memory, source memory, and guessing.

Manipulating one Source Dimension

Experiment 1 aimed at investigating whether source memory for item-context associations benefits from sleep by manipulating the spatial position (left vs. right) of pictures on a computer screen in a standard source-monitoring task (e.g., Bayen et al., 1996; Murnane & Bayen, 1996). To disentangle cognitive processes that are likely confounded in commonly used source memory measures, Submodel 5a of the two-high-threshold multinomial model of source monitoring (2HTSM; Bayen et al., 1996) was used. This model provides separate parameter estimates for item memory (parameter D), source memory (parameter d), and guessing (parameters a, b, and g; for details on the standard source-monitoring task and the 2HTSM, see Section 3). Specifically, the active systems consolidation account predicts better item memory (D) and source memory (d) in the sleep than in the wake condition for a 12-hr retention interval. This was tested in an online experiment ($N = 132$) by comparing participants randomly assigned either to a wake or a sleep condition.

Supporting the active systems consolidation account, the model-based results showed clear evidence for better source memory in the sleep condition. In contrast, the empirical evidence for a sleep benefit in item memory was mixed. This is, however, in line with previous research (e.g., Köster et al., 2017; Mawdsley et al., 2014; van der Helm et al., 2011; Wang & Fu, 2009). One possible explanation for this result is that sleep benefits for word materials appear to be relatively small in recognition tasks compared to cued or free recall tasks (Berres & Erdfelder, 2021). Because recognition tasks impose a much lower burden on memory retrieval than, for example, free recall tasks, small sleep benefits in item recognition might be difficult to detect. To conclude, although the results of Experiment 1 show that sleep improves source memory within a 12-hr retention interval as predicted by the active systems consolidation account, an experimental follow-up evaluation is required to establish the validity of this finding.

Manipulating two Source Dimensions

The main purpose of Experiment 2 was therefore to conceptually replicate the results for spatial position memory of Experiment 1. By manipulating frame color (blue vs. yellow) orthogonally to the spatial position (left vs. right) of pictures, Experiment 2 also addressed two additional research questions: First, it was investigated whether the results for spatial position memory can be generalized to other source dimensions (i.e., frame color). Second, it was examined whether sleep compared to wakefulness improves memory for context-context associations (i.e., bound source memory for spatial position and frame color). To disentangle the hypothesized latent processes involved, a reparameterized variant of the multinomial model of multidimensional source monitoring was employed (Meiser, 2014).

This model is based on an extended source-monitoring task (i.e., instead of one, two source dimensions—spatial position and frame color—are manipulated orthogonally) and provides parameter estimates for item memory (parameter D), bound source memory (i.e., spatial position plus frame color; parameter d), unbound source memory (parameters $e^{Position}$ and e^{Color}), and guessing (parameters $a^{Position}$, $a^{Color}_{|left}$, $a^{Color}_{|right}$, b, $g^{Position}$, $g^{Color}_{|left}$, and $g^{Color}_{|right}$; Meiser, 2014). According to the active systems consolidation account, item memory (D), bound source memory (i.e., context-context binding; d), and unbound source memory ($e^{Position}$, e^{Color}) should be larger when participants sleep during the 12-hr retention interval than when they stay awake. To test these predictions, an online experiment ($N = 134$) with a randomized 2×2 mixed factorial design, including source dimension (spatial position vs. frame color) as within-subject factor and wake versus sleep as between-subjects factor was conducted.

Replicating Experiment 1 and in line with the active systems consolidation account, the model-based results showed a statistically significant sleep benefit in unbound source memory (i.e., item-context binding) for source dimension "spatial position" ($e^{Position}$). However, in contrast to what was predicted, no statistically significant sleep benefit emerged in unbound source memory for source dimension "frame color" (e^{Color}) and in context-context binding (d). This is probably due to a floor effect in the encoding of the frame color context, indicating that a sufficiently high level of memory strength at encoding is necessary for the sleep benefit to occur (cf. Denis et al., 2020; Muehlroth et al., 2020; Rauchs et al., 2011). Finally, the empirical evidence for a sleep benefit in item memory was again mixed.

In sum, both experiments consistently showed a sleep benefit in source memory for spatial position in line with the active systems consolidation account. The results also suggest that relevant context features, such as spatial position or frame color, must be sufficiently salient and thus well encoded for the sleep benefit to occur. Further research with sufficiently high memory strength of item-context and context-context associations at encoding is therefore needed to clarify the impact of sleep on (a) source dimensions other than spatial position, and (b) context-context binding. Nevertheless, this article presents additional support for the active systems consolidation account incorporated in the integrative framework. In addition, it provides a further demonstration that MPT models are a powerful alternative to frequently used source memory measures.

In the following section, I first summarize the evidence for the integrative framework provided by my thesis before discussing limitations and directions for future research.

4 Discussion

Almost a century of research on the sleep benefit in episodic memory has yielded important insights into the neurocognitive mechanisms underlying the effect. Despite this progress, there is still a lot of controversy how to best explain the sleep benefit. Currently, a variety of theories about the underlying mechanisms exist that require further refinement and more rigorous testing. Moreover, the mechanisms proposed by these theories have commonly been discussed as mutually exclusive (e.g., "Does sleep improve episodic memory by boosting memory consolidation or by reducing retroactive interference?").

However, there are several findings that can hardly be explained by one theory in isolation without making additional assumptions. In my thesis, I therefore proposed an integrative, multicausal framework that provides a broader and more comprehensive explanation of the sleep benefit. Furthermore, using a multinomial modeling approach, I tested core assumptions of the theories incorporated in the integrative framework to gain a deeper understanding of the neurocognitive mechanisms involved. As such, the integrative framework was developed based on meta-analytic results which showed a robust sleep benefit of moderate size in episodic memory (Article 1; Berres & Erdfelder, 2021).

To reiterate, according to the integrative framework sleep compared to wakefulness simultaneously improves both episodic memory storage and retrieval by memory consolidation (active systems and opportunistic consolidation account) and reduction of retroactive interference (interference account), respectively. This prediction received support from two experiments by Erdfelder et al. (2022) who employed an appropriately designed MPT model to investigate how sleep compared to wakefulness affects episodic memory storage and retrieval.

4.1 Testing Core Assumptions of the Integrative Framework

Here, I adopted this multinomial modeling approach to test two core assumptions of the active systems consolidation account, which is incorporated in the integrative framework. These core assumptions were not sufficiently addressed in the meta-analysis because they require a more thorough investigation of the underlying mechanisms.

The first core assumption of the active systems consolidation account predicts that information salience in form of encoding strength and test expectation moderates sleep benefits in episodic memory storage. However, previous research suggests that encoding strength also affects interference-based sleep benefits in memory retrieval (cf. Howe, 1970; McGeoch, 1929). In their first experiment, Berres and Erdfelder (2023, Article 2) therefore addressed the question whether encoding strength affects the sleep benefit in episodic memory storage, retrieval, or both. This was investigated by employing the EMR model and manipulating encoding strength by presenting word-pairs once versus three times (i.e., medium vs. strong encoding).

The model-based results replicated Erdfelder et al.'s (2022) finding that sleep fosters both episodic memory storage and retrieval, providing additional support for the integrative framework. Furthermore, the results also showed that encoding strength affects the sleep benefit in episodic memory retrieval rather than storage. This finding is at odds with the active systems consolidation account but in line with independent research showing that retroactive-interference effects decrease with encoding strength.

Moreover, test expectation appears not to affect the sleep benefit in episodic memory storage which conflicts with the active systems consolidation account. Taken together, the current empirical evidence suggest that the interference account offers a better explanation of encoding-strength effects on the sleep benefit than the active systems consolidation account. More importantly, the results also indicate that sleep benefits in episodic memory are best explained by considering both the active systems consolidation and the interference account as suggested by the integrative framework.

The second core assumption of the active systems consolidation account predicts that sleep improves source memory for retention intervals of up to 12 hr (Inostroza & Born, 2013; Klinzing et al., 2019). However, previous research has shown mixed results for a sleep benefit in source memory. One possible reason for the inconsistent findings could be the use of source memory measures that confound item memory, source memory, and guessing. Berres et al. (2023, Article 3) therefore reexamined the question whether sleep benefits source memory with a multinomial modeling approach to provide uncontaminated measures of source memory. The model-based results showed a clear sleep benefit in source memory for sufficiently strong encoded spatial positions and thus support the active systems consolidation account incorporated in the integrative framework.

In sum, the three articles provide convincing evidence for the integrative framework. However, as detailed in the following section, further research is needed to comprehensively test the integrative framework and to generalize findings beyond the employed sleep study design of natural sleep and wakefulness (cf. Article 1; Berres & Erdfelder, 2021). As demonstrated in this thesis, MPT models are powerful tools to investigate the sleep benefit in episodic memory. Adopting the multinomial modeling approach and enriching it with physiological data from neuroimaging and polysomnography in future research is a promising way to expand our knowledge of the sleep benefit. Although this thesis contributes towards a better understanding of the sleep benefit in episodic memory, there are several limitations and open questions that are discussed in the following section.

4.2 Limitations and Directions for Future Research

By showing encouraging initial support for the integrative framework, this thesis sets the starting point for follow-up experiments to establish the truth status of the framework more rigorously. In what follows, I describe limitations and outline directions for future research by focusing on three important points that are not sufficiently addressed in the three articles of this thesis: (a) the evaluation of the opportunistic consolidation account, (b) the further validation of the EMR model, (c) the closer examination of encoding strength, and (d) the generalization of empirical evidence. Finally, I discuss theoretical, methodological, and practical contributions of the present thesis to sleep-and-memory research.

Evaluating the Opportunistic Consolidation Account

The major focus of my thesis is on evaluating the active systems consolidation and interference accounts. Although the circadian-effects account was considered as an alternative explanation and ruled out by previous research (see, for example, Erdfelder et al., 2022), the thesis remains relatively silent about the opportunistic consolidation account. This theoretical account predicts that episodic memory storage can be improved by any state that reduces retroactive interference and thus provides a beneficial environment for memory consolidation. Consequently, consolidation-based benefits in episodic memory storage should result not only from sleep but also from other states, such as quiet wake and consumption of alcohol or benzodiazepines (e.g., Ellenbogen et al., 2006; Mednick et al., 2011; Wixted, 2004, 2005).

To inform about benefits of sleep compared to quiet wake in memory performance, an exploratory analysis using multilevel meta-regression with RVE was conducted (see Article 1; Berres & Erdfelder, 2021). Descriptively, the sleep benefit relative to wakeful rest was even larger than relative to usual everyday activity and controlled cognitive activity during the whole or part of the retention interval. This suggests that physiological processes associated with sleep rather than reduced retroactive interference due to cognitive inactivity drive the sleep benefit. Thus, this finding contradicts the opportunistic consolidation account. However, the number of currently available effect sizes for wakeful rest is too small for a meaningful interpretation (Article 1; Berres & Erdfelder, 2021). Thus, future research is needed to test the opportunistic consolidation account experimentally by comparing sleep with low and high interference wake states while keeping the retention interval constant.

Moreover, in contrast to the active systems consolidation account, the opportunistic consolidation account states that neither sleep nor SWS possess specific features that are important for memory consolidation. Future research should therefore examine physiological data from neuroimaging and polysomnography in addition to behavioral data. Indeed, including physiological data (e.g., sleep spindle density) as covariates in MPT model-based analyses may prove particularly beneficial in gaining deeper insights into the neurocognitive mechanisms underlying the sleep benefit. Finally, as sketched above, the opportunistic consolidation account predicts that any state that reduces retroactive interference (e.g., quiet wake, alcohol, benzodiazepines) should boost episodic memory storage. To answer the question whether results from sleep versus quiet wake comparisons can be generalized to other states of reduced retroactive interference, more research is needed. Again, MPT models may provide valuable information about the underlying processes.

Consider, for example, the study by Quevedo Pütter and Erdfelder (2022) in which the EMR model was used to investigate the effect of alcohol consumption after learning on episodic memory storage (m) and retrieval (r_f). According to the opportunistic consolidation account, alcohol consumption should improve episodic memory storage (m) by reducing retroactive interference and thus creating a beneficial environment for memory consolidation. This hypothesis was tested in a lab experiment ($N = 93$) with a randomized between-subjects design, including an alcohol and a placebo condition. The authors found a statistically significant positive effect of alcohol consumption after learning on free recall retrieval (r_f) but not on episodic memory storage (m). Thus, these results contradict the opportunistic consolidation account.

Validating the Encoding-Maintenance-Retrieval Model

In Erdfelder et al. (2022), the validity of the EMR model was established by showing that experimental manipulations selectively affected specific model parameters (cf. Erdfelder et al., 2009; Hütter & Klauer, 2016). To elaborate, manipulating the length of retention intervals selectively affected storage (m) and retrieval (r_f) but not encoding (e; Experiment 1), whereas manipulating the presence of retrieval cues in free recall selectively affected retrieval (r_f) but neither encoding (e) nor storage (m; Experiment 2).

According to the active systems consolidation account, SWS is important for memory consolidation. In fact, Plihal and Born (1997) observed a sleep benefit for participants who slept the first half of the night, which is predominantly characterized by SWS.

In contrast, no sleep benefit emerged for participants who slept the second half of the night, which is predominantly characterized by REM sleep (Plihal & Born, 1997). Based on such a split-night design, the validation technique outlined above can help to further establish the EMR model's validity by showing that SWS boosts memory storage (i.e., parameter m). More specifically, episodic memory storage (m) should be larger after the first SWS-rich half of the night than after the second half, which is predominantly characterized by REM sleep. By contrast, encoding (e) and retrieval (r_f) should barely be affected by slow-wave-rich versus REM-rich sleep as predicted by the active systems consolidation account.

However, sleep study designs are often tailored to examine effects of specific sleep stages (see, for example, the split-night design), resulting in a confound between sleep parameters and sleep study designs. To check whether the results of split-night validation studies can be generalized to other sleep study designs, future research should investigate the discriminant validity of the EMR model's parameters by analyzing SWS-specific features as covariates (cf. Bott et al., 2020). Specifically, the active systems consolidation account predicts that SWS-specific features, such as the density of sleep spindles, should covary with episodic memory storage (m) but neither with encoding (e) nor retrieval (r_f). This prediction should hold for all sleep study designs that include a learning phase, a subsequent retention interval filled with either a period of sleep or wakefulness, and finally a testing phase. The combination of the two techniques thus provides a more rigorous measurement validation by testing whether the model parameters reliably capture the same neurocognitive mechanisms across different sleep study designs. This check is necessary to gain a better understanding of the mechanisms underlying memory enhancing sleep intervention techniques like targeted memory reactivation (TMR) by applying the EMR model (for details on TMR, see Feld & Diekelmann, 2020; Hu et al., 2020).

Examining Encoding-Strength Effects

In the first experiment of the second article in the present thesis (Berres & Erdfelder, 2023), the EMR model was employed to investigate the underlying mechanisms of encoding-strength effects on the sleep benefit in episodic memory. Focusing on the range of medium to high memory strength, encoding strength was manipulated by presenting half of the stimuli once versus three times. Thus, as in most previous studies, only two levels of memory strength were compared. However, a more rigorous test of Stickgold's (2009) assumption that the sleep benefit follows an inverted U-shaped function of memory strength is needed. Future research should therefore systematically vary encoding strength in small steps over a wide range, ideally using the same encoding strength manipulation.

In addition, further research is needed to better understand how various manipulations of information salience affect the sleep benefit and its underlying processes. As such, information salience can be manipulated not only by varying the number of stimulus presentations or other learning procedures (e.g., Abel et al., 2019; Bäuml et al., 2014; Bennion et al., 2016; Denis et al., 2020) but also by varying the emotionality of the stimuli (e.g., Hu et al., 2006; Payne et al., 2015). Because emotional stimuli receive more attention and thus more elaboration than neutral stimuli at encoding (Hamann, 2001), it is assumed that emotional stimuli profit more from memory consolidation during sleep (e.g., Payne & Kensinger, 2010; Walker & van der Helm, 2009).

Although more research is needed to test this prediction experimentally, the first article of this thesis provides tentative support by showing a descriptively larger sleep benefit for negative compared to positive and neutral stimuli (see Berres & Erdfelder, 2021, p. 1333). Specifically, in the first article of this thesis, emotionality, among a wide range of potential moderator variables, such as age, memory tests, stimulus materials, or total sleep time, was examined (see Berres & Erdfelder, 2021, p. 1314). However, for some potential moderator variables the available sample of effect sizes was simply too small, or the moderators contained too many missing values to be included in a multiple-moderator analysis. Thus, it was not possible to address all moderator variables of interest (e.g., motivational aspects like reward; Diekelmann et al., 2009; Stickgold & Walker, 2013).

In addition, for some of the analyzed moderator variables and their levels, the available number of effect sizes was relatively small which introduced uncertainty in the respective analysis (see, for example, the subgroup moderator analyses for emotionality or the exploratory moderator analysis for the opportunistic consolidation account outlined above). Consequently, nonsignificant results in moderator analyses, such as for emotionality, should be considered with caution. Further research is needed to investigate these potential moderators experimentally.

Moreover, future research should investigate whether effects of emotionality on sleep benefits in episodic memory storage also follow an inverted U-shaped curve, as predicted by the active systems consolidation account for effects of encoding strength. Here, the EMR model is a useful method to examine whether and how the cognitive processes underlying the sleep benefit in episodic memory are affected by emotionality, leading to a better understanding of both emotionality and encoding-strength effects on the sleep benefit.

In the third article (Berres et al., 2023), the question whether sleep compared to wakefulness benefits source memory for 12-hr retention intervals was examined. Contradicting the active systems consolidation account, no significant sleep benefit emerged for item-context binding for source dimension “frame color” and context-context binding, probably due to floor effects. Thus, the prediction of the active systems consolidation account should be reinvestigated in future research, ensuring sufficiently high encoding strength, for example, by intentional learning of item-context associations. More research is also needed to determine the precise time course of memory consolidation during sleep, which is not yet well understood (e.g., Klinzing et al., 2019; Lewis & Durrant, 2011; Pöhlchen & Schönauer, 2020; Stickgold & Walker, 2007). This point can be addressed by conducting experiments with standardized study designs and varying retention intervals to determine the minimum and maximum retention interval length for which sleep benefits in source memory occur.

Generalizing Empirical Evidence

Finally, compared to online experiments (cf. Article 3), lab experiments (cf. Article 2) are characterized by small sample sizes, which are typical for sleep and memory research due to time and cost intensive study designs and methods (Berres & Erdfelder, 2021). In addition, as in this thesis, most participants are healthy young adults who belong to western, educated, industrialized, rich, and democratic (WEIRD) populations, making it difficult to generalize the findings (cf. Falk et al., 2013; Henrich et al., 2010). However, the collection of larger and more diverse samples could be facilitated by supporting collaborations between institutions (e.g., clinics, universities) or by conducting online studies—if the research question and the study design permits.

As sketched above, this thesis focused on healthy human individuals and is thus silent about clinical populations. Investigating the sleep benefit in clinical populations and comparing the results to a healthy control group can, however, contribute to a better understanding of the underlying mechanisms. For example, previous studies have found a lower sleep benefit in children with attention deficit hyperactivity disorder (ADHD) compared to typically developing children (e.g., Prehn-Kristensen et al., 2011; Prehn-Kristensen et al., 2017). Here, the EMR model is a prime candidate to clarify the processes that lead to a reduced sleep benefit in episodic memory among children with ADHD, providing valuable information for theory building and memory-enhancing interventions.

In addition, most of the participants investigated in this thesis are young adults. Because sleep benefits may decrease across the life span due to age-related changes in sleep parameters, further research is needed to check whether the results can be generalized to other age groups (for more details, see Berres & Erdfelder, 2021). Again, MPT models are well suited—particularly because of the capability to analyze sleep parameters as covariates—to tackle the question whether and how age affects the sleep benefit and its underlying mechanisms.

4.3 Contributions to Sleep-and-Memory Research

Despite the outlined limitations, the present thesis makes significant theoretical, methodological, and practical contributions to the research on sleep and memory: First, by systematically investigating a wide range of potential moderator variables and by introducing the integrative framework, this thesis makes an important theoretical contribution. As such, knowledge about variables that moderate the sleep benefit is of prime importance as they offer a better understand of the processes underlying the sleep benefit. Furthermore, the knowledge about moderators also allows to test corresponding theories, to explain unexpected results, and to design informative replication studies. In fact, across the three articles in the present thesis, encoding strength appears to be an important moderator of the sleep benefit that should be carefully considered when designing studies and interpreting results. The results of the systematic investigation of moderators in the meta-analysis was also used to develop the integrative framework.

Although a single theory that coherently explains why sleep compared to wakefulness improves both episodic memory storage and retrieval would be an attractive alternative, the currently available evidence supports the integrative framework. For example, Diekelmann et al. (2009) suggested that systems consolidation (i.e., synaptic changes across various brain areas) during sleep improves memory retrieval in cued and free recall tasks. Because memory representations are integrated into preexisting memory networks, it is assumed that they are linked to a larger number of cues that facilitate access and thus memory retrieval. However, this assumption is challenged by empirical evidence showing that no significant behavioral effects of systems consolidation emerge within 10 to 24 hr after learning (e.g., Lutz et al., 2017; Takashima et al., 2006). For this reason, the integrative framework offers a more plausible explanation of the sleep benefit for retention intervals up to 12 hr.

Second, the present thesis also makes an important methodological contribution by demonstrating that MPT models are powerful tools to investigate sleep benefits in episodic memory more thoroughly. Specifically, by using appropriately designed MPT models, Article 2 (Berres & Erdfelder, 2023) showed that variations of the sleep benefit in episodic memory may not always be due to variations in storage but also be due to variations in retrieval success. Also applying multinomial modeling, Article 3 (Berres et al., 2023) provided additional evidence that memory consolidation is one of the key processes that contributes to the sleep benefit in episodic memory.

Third, by quantifying the sleep benefit in episodic memory, this thesis makes a contribution of practical significance. As such, knowledge about the size of the effect is essential when planning studies on the sleep benefit that are sufficiently powered. Previously, researchers had to rely on reported effect sizes from single experiments and on their gut feeling in order to ensure adequately powered studies. The meta-analysis included in this thesis now informs about effect sizes under a variety of context conditions. In this way, the present work helps to plan future sleep-and-memory studies appropriately and to increase replicability of research on the sleep benefit in episodic memory.

Overall, this thesis contributes to a better understanding of the sleep benefit in episodic memory, demonstrates the power of MPT models in investigating the underlying mechanisms, and provides guidance to increase the replicability of the sleep benefit in future research.

5 Conclusion

Even though the sleep benefit in episodic memory is considered a well-established finding, knowledge about the typical size of the effect and its underlying mechanisms was still limited. In my thesis, I have contributed to fill this gap by quantifying the sleep benefit—which indeed proved robust—and by promoting a theoretical framework that integrates prominent theories from neuroscience and psychology. As such, this integrative framework was developed based on meta-analytic results and integrates active systems consolidation, opportunistic consolidation, and interference accounts. It received initial empirical support from two articles in which core assumptions of the incorporated theories were tested with appropriately designed MPT models. Overall, both the integrative framework and the multinomial modeling approach promoted in this thesis have the potential to advance our understanding of the underlying neurocognitive mechanisms in the future. Thereby, they may prompt the development of effective memory-enhancing sleep interventions to maintain proper episodic memory functioning, especially in clinical populations.

6 References

Abel, M., & Bäuml, K.-H. T. (2012). Retrieval-induced forgetting, delay, and sleep. *Memory*, *20*(5), 420–428. https://doi.org/10.1080/09658211.2012.671832

Abel, M., & Bäuml, K.-H. T. (2013a). Adaptive memory: The influence of sleep and wake delay on the survival-processing effect. *Journal of Cognitive Psychology*, *25*(8), 917–924. https://doi.org/10.1080/20445911.2013.825621

Abel, M., & Bäuml, K.-H. T. (2013b). Sleep can eliminate list-method directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 946–952. https://doi.org/10.1037/a0030529

Abel, M., Haller, V., Köck, H., Pötschke, S., Heib, D., Schabus, M., & Bäuml, K.-H. T. (2019). Sleep reduces the testing effect—But not after corrective feedback and prolonged retention interval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(2), 272–287. https://doi.org/10.1037/xlm0000576

Ambrosini, M. V., & Giuditta, A. (2001). Learning and sleep: The sequential hypothesis. *Sleep Medicine Reviews*, *5*(6), 477–490. https://doi.org/10.1053/smrv.2001.0180

Antony, J. W., & Paller, K. A. (2017). Hippocampal contributions to declarative memory consolidation during sleep. In D. F. Hannula & M. C. Duff (Eds.), *The hippocampus from cells to systems* (1st ed., pp. 245–280). Springer International Publishing. https://doi.org/10.1007/978-3-319-50406-3_9

Antony, J. W., & Schapiro, A. C. (2019). Active and effective replay: Systems consolidation reconsidered again. *Nature Reviews Neuroscience*, *20*, 506–507. https://doi.org/10.1038/s41583-019-0191-8

Ashton, J. E., & Cairney, S. A. (2021). Future-relevant memories are not selectively strengthened during sleep. *PLOS ONE*, *16*(11), Article e0258110. https://doi.org/10.1371/journal.pone.0258110

Backhaus, J., & Junghanns, K. (2006). Daytime naps improve procedural motor memory. *Sleep Medicine*, *7*(6), 508–512. https://doi.org/10.1016/j.sleep.2006.04.002

Bäuml, K.-H. T., Holterman, C., & Abel, M. (2014). Sleep can reduce the testing effect: It enhances recall of restudied items but can leave recall of retrieved items unaffected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1568–1581. https://doi.org/10.1037/xlm0000025

Barco, A., Lopez de Armentia, M., & Alarcon, J. M. (2008). Synapse-specific stabilization of plasticity processes: The synaptic tagging and capture hypothesis revisited 10 years later. *Neurscience and Biobehavioral Reviews*, *32*(4), 831–851. https://doi.org/10.1016/j.neubiorev.2008.01.002

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, *97*(4), 548–564. https://doi.org/10.1037/0033-295X.97.4.548

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*(1), 57–86. https://doi.org/10.3758/BF03210812

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 197–215. https://doi.org/10.1037/0278-7393.22.1.197

Bennion, K. A., Payne, J. D., & Kensinger, E. A. (2016). The impact of napping on memory for future-relevant stimuli: Prioritization among multiple salience cues. *Behavioral Neuroscience*, *130*(3), 281–289. https://doi.org/10.1037/bne0000142

Berres, S., & Erdfelder, E. (2021). The sleep benefit in episodic memory: An integrative review and a meta-analysis. *Psychological Bulletin*, *147*(12), 1309–1353. https://doi.org/10.1037/bul0000350

Berres, S., & Erdfelder, E. (2023). *Is it all about storage? Effects of encoding strength and test expectation on the sleep benefit in episodic memory* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

Berres, S., Erdfelder, E., & Kuhlmann, B. G. (2023). *Does sleep benefit source memory? Investigating 12-hr retention intervals with a multinomial modeling approach* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

Blackman, J., Swirski, M., Clynes, J., Harding, S., Leng, Y., & Coulthard, E. (2021). Pharmacological and non-pharmacological interventions to enhance sleep in mild cognitive impairment and mild Alzheimer's disease: A systematic review. *Journal of Sleep Research*, *30*(4), Article e13229. https://doi.org/10.1111/jsr.13229

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

Bott, F. M., Heck, D. W., & Meiser, T. (2020). Parameter validation in hierarchical MPT models by functional dissociation with continuous covariates: An application to contingency inference. *Journal of Mathematical Psychology*, *98*, Article 102388. https://doi.org/10.1016/j.jmp.2020.102388

Bröder, A., & Meiser, T. (2007). Measuring source memory. *Zeitschrift für Psychologie / Journal of Psychology*, *215*(1), 52–60. https://doi.org/10.1027/0044-3409.215.1.52

Cirelli, C., & Tononi, G. (2015). Sleep and synaptic homeostasis. *Sleep*, *38*(1), 161–162. https://doi.org/10.5665/sleep.4348

Cordi, M. J., & Rasch, B. (2021). How robust are sleep-mediated memory benefits? *Current Opinion in Neurobiology*, *67*, 1–7. https://doi.org/10.1016/j.conb.2020.06.002

Denis, D., Mylonas, D., Poskanzer, C., Bursal, V., Payne, J. D., & Stickgold, R. (2021). Sleep spindles preferentially consolidate weakly encoded memories. *Journal of Neuroscience*, *41*(18), 4088–4099. https://doi.org/10.1523/JNEUROSCI.0818-20.2021

Denis, D., Schapiro, A. C., Poskanzer, C., Bursal, V., Charon, L., Morgan, A., & Stickgold, R. (2020). The role of item exposure and visualization success in the consolidation of memories across wake and sleep. *Learning & Memory*, *27*, 451–456. https://doi.org/10.1101/lm.051383.120

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

Diekelmann, S., Wilhelm, I., & Born, J. (2009). The whats and whens of sleep-dependent memory consolidation. *Sleep Medicine Reviews*, *13*(5), 309–321. https://doi.org/10.1016/j.smrv.2008.08.002

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annual Review of Psychology*, *55*, 51–86. https://doi.org/10.1146/annurev.psych.55.090902.142050

Dudai, Y. (2012). The restless engram: Consolidations never end. *Annual Review of Neuroscience*, *35*, 227–247. https://doi.org/10.1146/annurev-neuro-062111-150500

Dudai, Y., Karni, A., & Born, J. (2015). The consolidation and transformation of memory. *Neuron*, *88*(1), 20–32. https://doi.org/10.1016/j.neuron.2015.09.004

Dyne, A. M., Humphreys, M. S., Bain, J. D., & Pike, R. (1990). Associative interference effects in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(5), 813–824. https://doi.org/10.1037/0278-7393.16.5.813

Ebbinghaus, H. (1885). *Über das Gedächtnis* [On memory] (1st ed.). German Text Archive. http://www.deutschestextarchiv.de/ebbinghaus_gedaechtnis_1885

Ecker, U. K. H., Brown, G. D. A., & Lewandowsky, S. (2015). Memory without consolidation: Temporal distinctiveness explains retroactive interference. *Cognitive Science*, *39*(7), 1570–1593. https://doi.org/10.1111/cogs.12214

Ecker, U. K. H., Tay, J.-X., & Brown, G. D. A. (2015). Effects of prestudy and poststudy rest on memory: Support for temporal interference accounts of forgetting. *Psychonomic Bulletin & Review*, *22*, 772–778. https://doi.org/10.3758/s13423-014-0737-8

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, *30*, 123–152. https://doi.org/10.1146/annurev.neuro.30.051606.094328

Ekstrand, B. R. (1967). Effect of sleep on memory. *Journal of Experimental Psychology*, *75*(1), 64–72. https://doi.org/10.1037/h0024907

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. https://doi.org/10.1016/j.conb.2006.10.006

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie*, *217*(3), 108–124. https://doi.org/10.1027/0044-3409.217.3.108

Erdfelder, E., Berres, S., Quevedo Pütter, J., & Küpper-Tetzel, C. E. (2022). *Why does sleep improve episodic memory? An encoding-maintenance-retrieval analysis* [Manuscript under revision]. Department of Psychology, University of Mannheim.

Falk, E. B., Hyde, L. W., Mitchell, C., Faul, J., Gonzalez, R., Heitzeg, M. M., Keating, D. P., Langa, K. M., Martz, M. E., Maslowsky, J., Morrison, F. J., Noll, D. C., Patrick, M. E., Pfeffer, F. T., Reuter-Lorenz, P. A., Thomason, M. E., Davis-Kean, P., Monk, C. S., & Schulenberg, J. (2013). What is a representative brain? Neuroscience meets population science. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(44), 17615–17622. https://doi.org/10.1073/pnas.1310134110

Feld, G. B., & Born, J. (2017). Sculpting memory during sleep: Concurrent consolidation and forgetting. *Current Opinion in Neurobiology*, *44*, 20–27. https://doi.org/10.1016/j.conb.2017.02.012

Feld, G. B., & Diekelmann, S. (2020). Building the bridge: Outlining steps toward an applied sleep-and-memory research program. *Current Directions in Psychological Science*, *29*(6), 554–562. https://doi.org/10.1177/0963721420964171

Feld, G. B., Lange, T., Gais, S., & Born, J. (2013). Sleep-dependent declarative memory consolidation—Unaffected after blocking NMDA or AMPA receptors but enhanced by NMDA coagonist D-cycloserine. *Neuropsychopharmacology*, *38*, 2688–2697. https://doi.org/10.1038/npp.2013.179

Fenn, K. M., & Hambrick, D. Z. (2012). Individual differences in working memory capacity predict sleep-dependent memory consolidation. *Journal of Experimental Psychology: General*, *141*(3), 404–410. https://doi.org/10.1037/a0025268

Fenn, K. M., & Hambrick, D. Z. (2013). What drives sleep-dependent memory consolidation: Greater gain or less loss? *Psychonomic Bulletin & Review*, *20*, 501–506. https://doi.org/10.3758/s13423-012-0366-z

Frey, U., & Morris, R. G. M. (1998). Synaptic tagging: Implications for late maintenance of hippocampal long-term potentiation. *Trends in Neurosciences*, *21*(5), 181–188. https://doi.org/10.1016/S0166-2236(97)01189-2

Gais, S., & Born, J. (2004). Low acetylcholine during slow-wave sleep is critical for declarative memory consolidation. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(7), 2140–2144. https://doi.org/10.1073/pnas.0305404101

Giuditta, A. (2014). Sleep memory processing: The sequential hypothesis. *Frontiers in Systems Neuroscience*, *8*, Article 219. https://doi.org/10.3389/fnsys.2014.00219

Giuditta, A., Ambrosini, M. V., Montagnese, P., Mandile, P., Cotugno, M., Grassi Zucconi, G., & Vescia, S. (1995). The sequential hypothesis of the function of sleep. *Behavioural Brain Research*, *69*(1–2), 157–166. https://doi.org/10.1016/0166-4328(95)00012-I

Goerke, M., Müller, N. G., & Cohrs, S. (2017). Sleep-dependent memory consolidation and its implications for psychiatry. *Journal of Neural Transmission*, *124*, 163–178. https://doi.org/10.1007/s00702-015-1476-3

Gui, W.-J., Li, H.-J., Guo, Y.-H., Peng, P., Lei, X., & Yu, J. (2017). Age-related differences in sleep-based memory consolidation: A meta-analysis. *Neuropsychologia*, *97*, 46–55. https://doi.org/10.1016/j.neuropsychologia.2017.02.001

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Sciences*, *5*(9), 394–400. https://doi.org/10.1016/S1364-6613(00)01707-1

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, *50*, 264–284. https://doi.org/10.3758/s13428-017-0869-7

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Howe, T. S. (1970). Joint effects of proactive and retroactive interference as a function of degree of learning. *Journal of Experimental Psychology*, *83*(1), 68–72. https://doi.org/10.1037/h0028553

Hu, P., Stylos-Allan, M., & Walker, M. P. (2006). Sleep facilitates consolidation of emotional declarative memory. *Psychological Science*, *17*(10), 891–898. https://doi.org/10.1111/j.1467-9280.2006.01799.x

Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, *59*(1), 21–47. https://doi.org/10.1007/BF02294263

Hu, X., Cheng, L. Y., Chiu, M. H., & Paller, K. A. (2020). Promoting memory consolidation during sleep: A meta-analysis of targeted memory reactivation. *Psychological Bulletin*, *146*(3), 218–244. https://doi.org/10.1037/bul0000223

Hütter, M., & Klauer, K. C. (2016). Applying processing trees in social psychology. *European Review of Social Psychology*, *27*(1), 116–159. https://doi.org/10.1080/10463283.2016.1212966

Inostroza, M., & Born, J. (2013). Sleep for preserving and transforming episodic memory. *Annual Review of Neuroscience*, *36*, 79–102. https://doi.org/10.1146/annurev-neuro-062012-170429

Jenkins, J. G., & Dallenbach, K. M. (1924). Obliviscence during sleep and waking. *The American Journal of Psychology*, *35*(4), 605–612. https://doi.org/10.2307/1414040

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28. https://doi.org/10.1037/0033-2909.114.1.3

Jurewicz, K., Cordi, M. J., Staudigl, T., & Rasch, B. (2016). No evidence for memory decontextualization across one night of sleep. *Frontiers in Human Neuroscience*, *10*, Article 7. https://doi.org/10.3389/fnhum.2016.00007

Klinzing, J. G., Niethard, N., & Born, J. (2019). Mechanisms of systems memory consolidation during sleep. *Nature Neuroscience*, *22*(10), 1598–1610. https://doi.org/10.1038/s41593-019-0467-3

Köster, M., Finger, H., Kater, M.-J., Schenk, C., & Gruber, T. (2017). Neuronal oscillations indicate sleep-dependent changes in the cortical memory trace. *Journal of Cognitive Neuroscience*, *29*(4), 698–707. https://doi.org/10.1162/jocn_a_01071

Küpper-Tetzel, C. E., & Erdfelder, E. (2012). Encoding, maintenance, and retrieval processes in the lag effect: A multinomial processing tree analysis. *Memory*, *20*(1), 37–47. https://doi.org/10.1080/09658211.2011.631550

Küpper-Tetzel, C. E., Erdfelder, E., & Dickhäuser, O. (2014). The lag effect in secondary school classrooms: Enhancing students' memory for vocabulary. *Instructional Science*, *42*, 373–388. https://doi.org/10.1007/s11251-013-9285-2

Kuhlmann, B. G., Erdfelder, E., & Moshagen, M. (2019). Testing interactions in multinomial processing tree models. *Frontiers in Psychology*, *10*, Article 2364. https://doi.org/10.3389/fpsyg.2019.02364

Kuhlmann, B. G., Symeonidou, N., Tanyas, H., & Wulff, L. (2021). Chapter three - Remembering and reconstructing episodic context: An overview of source monitoring methods and behavioral findings. In K. D. Federmeier & L. Sahakyan (Eds.), *Psychology of learning and motivation* (pp. 79–124). Academic Press. https://doi.org/10.1016/bs.plm.2021.06.002

Leong, R. L. F., Cheng, G. H.-L., Chee, M. W. L., & Lo, J. C. (2019). The effects of sleep on prospective memory: A systematic review and meta-analysis. *Sleep Medicine Reviews*, *47*, 18–27. https://doi.org/10.1016/j.smrv.2019.05.006

Lewin, K. (1952). *Field theory in social science: Selected theoretical papers*. Tavistock.

Lewis, P. A., Cairney, S., Manning, L., & Critchley, H. D. (2011). The impact of overnight consolidation upon memory for emotional and neutral encoding contexts. *Neuropsychologia*, *49*(9), 2619–2629. https://doi.org/10.1016/j.neuropsychologia.2011.05.009

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

Lo, J. C., Groeger, J. A., Cheng, G. H., Dijk, D.-J., & Chee, M. W. L. (2016). Self-reported sleep duration and cognitive performance in older adults: A systematic review and meta-analysis. *Sleep Medicine*, *17*, 87–98. https://doi.org/10.1016/j.sleep.2015.08.021

Lutz, N. D., Diekelmann, S., Hinse-Stern, P., Born, J., & Rauss, K. (2017). Sleep supports the slow abstraction of gist from visual perceptual memories. *Scientific Reports*, *7*, Article 42950. https://doi.org/10.1038/srep42950

Mander, B. A., Winer, J. R., & Walker, M. P. (2017). Sleep and human aging. *Neuron*, *94*(1), 19–36. https://doi.org/10.1016/j.neuron.2017.02.004

Marshall, L., Mölle, M., Hallschmid, M., & Born, J. (2004). Transcranial direct current stimulation during sleep improves declarative memory. *Journal of Neuroscience*, *24*(44), 9985–9992. https://doi.org/10.1523/JNEUROSCI.2725-04.2004

Mawdsley, M., Grasby, K., & Talk, A. (2014). The effect of sleep on item recognition and source memory recollection among shift-workers and permanent day-workers. *Journal of Sleep Research*, *23*(5), 538–544. https://doi.org/10.1111/jsr.12149

McGeoch, J. A. (1929). The influence of degree of learning upon retroactive inhibition. *The American Journal of Psychology*, *41*(2), 252–262. https://doi.org/10.2307/1415236

McKinney, F. (1935). Retroactive inhibition and recognition memory. *Journal of Experimental Psychology*, *18*(5), 585–598. https://doi.org/10.1037/h0054474

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*(10), 504–514. https://doi.org/10.1016/j.tins.2011.06.003

Meiser, T. (2014). Analyzing stochastic dependence of cognitive processes in multidimensional source recognition. *Experimental Psychology*, *61*(5), 402–415. https://doi.org/10.1027/1618-3169/a000261

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neuronal mechanisms of source memory? *Psychological Bulletin*, *135*(4), 638–677. https://doi.org/10.1037/a0015849

Morris, R. G. M. (2006). Elements of a neurobiological theory of hippocampal function: The role of synaptic plasticity, synaptic tagging and schemas. *European Journal* *of Neuroscience*, *23*(11), 2829–2846. https://doi.org/10.1111/j.1460-9568.2006.04888.x

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*(1), 42–54. https://doi.org/10.3758/BRM.42.1.42

Muehlroth, B. E., Rasch, B., & Werkle-Bergner, M. (2020). Episodic memory consolidation during sleep in healthy aging. *Sleep Medicine Reviews*, *52*, Article 101304. https://doi.org/10.1016/j.smrv.2020.101304

Murayama, K., Miyatsu, T., Buchli, D., & Storm, B. C. (2014). Forgetting as a consequence of retrieval: A meta-analytic review of retrieval-induced forgetting. *Psychological Bulletin*, *140*(5), 1383–1409. https://doi.org/10.1037/a0037505

Murnane, K., & Bayen, U. J. (1996). An evaluation of empirical measures of source identification. *Memory & Cognition*, *24*(4), 417–428. https://doi.org/10.3758/BF03200931

O'Neill, J., Pleydell-Bouverie, B., Dupret, D., & Csicsvari, J. (2010). Play it again: Reactivation of waking experience and memory. *Trends in Neurosciences*, *33*(5), 220–229. https://doi.org/10.1016/j.tins.2010.01.006

Pace-Schott, E. F., Germain, A., & Milad, M. R. (2015a). Effects of sleep on memory for conditioned fear and fear extinction. *Psychological Bulletin*, *141*(4), 835–857. https://doi.org/10.1037/bul0000014

Pace-Schott, E. F., Germain, A., & Milad, M. R. (2015b). Sleep and REM sleep disturbance in the pathophysiology of PTSD: The role of extinction memory. *Biology of Mood & Anxiety Disorders*, *5*, Article 3. https://doi.org/10.1186/s13587-015-0018-9

Palagini, L., Baglioni, C., Ciapparelli, A., Gemignani, A., & Riemann, D. (2013). REM sleep dysregulation in depression: State of the art. *Sleep Medicine Reviews*, *17*(5), 377–390. https://doi.org/10.1016/j.smrv.2012.11.001

Pan, S. C., & Rickard, T. C. (2015). Sleep and motor learning: Is there room for consolidation? *Psychological Bulletin*, *141*(4), 812–834. https://doi.org/10.1037/bul0000009

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. https://doi.org/10.1177/0963721410383978

Payne, J. D., Kensinger, E. A., Wamsley, E. J., Spreng, R. N., Alger, S. E., Gibler, K., Schacter, D. L., & Stickgold, R. (2015). Napping and the selective consolidation of negative aspects of scenes. *Emotion*, *15*(2), 176–186. https://doi.org/10.1037/a0038683

Payne, J. D., Stickgold, R., Swanberg, K., & Kensinger, E. A. (2008). Sleep preferentially enhances memory for emotional components of scenes. *Psychological Science*, *19*(8), 781–788. https://doi.org/10.1111/j.1467-9280.2008.02157.x

Petzka, M., Charest, I., Balanos, G. M., & Staresina, B. P. (2021). Does sleep-dependent consolidation favour weak memories? *Cortex*, *134*, 65–75. https://doi.org/10.1016/j.cortex.2020.10.005

Pfeiffer, B. E. (2020). The content of hippocampal “replay”. *Hippocampus*, *30*(1), 6–18. https://doi.org/10.1002/hipo.22824

Plihal, W., & Born, J. (1997). Effects of early and late nocturnal sleep on declarative and procedural memory. *Journal of Cognitive Neuroscience*, *9*(4), 534–547. https://doi.org/10.1162/jocn.1997.9.4.534

Pöhlchen, D., & Schönauer, M. (2020). Sleep-dependent memory consolidation in the light of rapid neocortical plasticity. *Current Opinion in Behavioral Sciences*, *33*, 118–125. https://doi.org/10.1016/j.cobeha.2020.02.001

Postman, L. (1952). Retroactive inhibition in recall and recognition. *Journal of Experimental Psychology*, *44*(3), 165–169. https://doi.org/10.1037/h0057388

Prehn-Kristensen, A., Göder, R., Fischer, J., Wilhelm, I., Seeck-Hirschner, M., Aldenhoff, J., & Baving, L. (2011). Reduced sleep-associated consolidation of declarative memory in attention-deficit/hyperactivity disorder. *Sleep Medicine*, *12*(7), 672–679. https://doi.org/10.1016/j.sleep.2010.10.010

Prehn-Kristensen, A., Molzow, I., Förster, A., Siebenhühner, N., Gesch, M., Wiesner, C. D., & Baving, L. (2017). Memory consolidation of socially relevant stimuli during sleep in healthy children and children with attention-deficit/hyperactivity disorder and oppositional defiant disorder: What you can see in their eyes. *Biological Psychology*, *123*, 196–204. https://doi.org/10.1016/j.biopsycho.2016.12.017

Pustejovsky, J. E., & Rodgers, M. A. (2019). Testing for funnel plot asymmetry of standardized mean differences. *Research Synthesis Methods*, *10*(1), 57–71. https://doi.org/10.1002/jrsm.1332

Quevedo Pütter, J., & Erdfelder, E. (2022). Alcohol-induced retrograde facilitation? Mixed evidence in a preregistered replication and encoding-maintenance-retrieval analysis. *Experimental Psychology*, *69*(6), 335–350. https://doi.org/10.1027/1618-3169/a000569

Rasch, B., & Born, J. (2013). About sleep’s role in memory. *Physiological Reviews*, *93*(2), 681–766. https://doi.org/10.1152/physrev.00032.2012

Rauchs, G., Feyers, D., Landeau, B., Bastin, C., Luxen, A., Maquet, P., & Collette, F. (2011). Sleep contributes to the strengthening of some memories over others, depending on hippocampal activity at learning. *Journal of Neuroscience*, *31*(7), 2563–2568. https://doi.org/10.1523/JNEUROSCI.3972-10.2011

Redondo, R. L., & Morris, R. G. M. (2011). Making memories last: The synaptic tagging and capture hypothesis. *Nature Reviews Neuroscience*, *12*, 17–30. https://doi.org/10.1038/nrn2963

Reverberi, S., Kohn, N., & Fernández, G. (2020). No evidence for an effect of explicit relevance instruction on consolidation of associative memories. *Neuropsychologia*, *143*, Article 107491. https://doi.org/10.1016/j.neuropsychologia.2020.107491

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*(3), 318–339. https://doi.org/10.1037/0033-295X.95.3.318

Rodgers, M. A., & Pustejovsky, J. E. (2020). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods*, *26*(2), 141–160. https://doi.org/10.1037/met0000300

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Schäfer, S. K., Wirth, B. E., Staginnus, M., Becker, N., Michael, T., & Sopp, M. R. (2020). Sleep's impact on emotional recognition memory: A meta-analysis of whole-night, nap, and rem sleep effects. *Sleep Medicine Reviews*, *51*, Article 101280. https://doi.org/10.1016/j.smrv.2020.101280

Schmid, D., Erlacher, D., Klostermann, A., Kredel, R., & Hossner, E.-J. (2020). Sleep-dependent motor memory consolidation in healthy adults: A meta-analysis. *Neuroscience and Biobehavioral Reviews*, *118*, 270–281. https://doi.org/10.1016/j.neubiorev.2020.07.028

Scullin, M. K. (2017). Do older adults need sleep? A review of neuroimaging, sleep, and aging studies. *Current Sleep Medicine Reports*, *3*, 204–214. https://doi.org/10.1007/s40675-017-0086-z

Scullin, M. K., & Bliwise, D. L. (2015). Sleep, cognition, and normal aging: Integrating a half century of multidisciplinary research. *Perspectives on Psychological Science*, *10*(1), 97–137. https://doi.org/10.1177/1745691614556680

Stickgold, R. (2009). How do i remember? Let me count the ways. *Sleep Medicine Reviews*, *13*(5), 305–308. https://doi.org/10.1016/j.smrv.2009.05.004

Stickgold, R., & Walker, M. P. (2007). Sleep-dependent memory consolidation and reconsolidation. *Sleep Medicine*, *8*(4), 331–343. https://doi.org/10.1016/j.sleep.2007.03.011

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

Takashima, A., Petersson, K. M., Rutters, F., Tendolkar, I., Jensen, O., Zwarts, M. J., McNaughton, B. L., & Fernández, G. (2006). Declarative memory consolidation in humans: A prospective functional magnetic resonance imaging study. *Proceed-*

ings of the National Academy of Sciences of the United States of America, *103*(3), 756–761. https://doi.org/10.1073/pnas.0507774103

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

Tononi, G., & Cirelli, C. (2006). Sleep function and synaptic homeostasis. *Sleep Medicine Reviews*, *10*(1), 49–62. https://doi.org/10.1016/j.smrv.2005.05.002

Tononi, G., & Cirelli, C. (2014). Sleep and the price of plasticity: From synaptic and cellular homeostasis to memory consolidation and integration. *Neuron*, *81*(1), 12–34. https://doi.org/10.1016/j.neuron.2013.12.025

Tononi, G., & Cirelli, C. (2020). Sleep and synaptic down-selection. *European Journal of Neuroscience*, *51*(1), 413–421. https://doi.org/10.1111/ejn.14335

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*, 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114

Tulving, E., & Psotka, J. (1971). Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. *Journal of Experimental Psychology*, *87*(1), 1–8. https://doi.org/10.1037/h0030185

van der Helm, E., Gujar, N., Nishida, M., & Walker, M. P. (2011). Sleep-dependent facilitation of episodic memory details. *PLOS ONE*, *6*(11), Article e27421. https://doi.org/10.1371/journal.pone.0027421

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect sizes in the presence of publication bias. *Psychometrika*, *60*(3), 419–435. https://doi.org/10.1007/BF02294384

Walker, M. P., & van der Helm, E. (2009). Overnight therapy? The role of sleep in emotional brain processing. *Psychological Bulletin*, *135*(5), 731–748. https://doi.org/10.1037/a0016570

Wang, B., & Fu, X.-I. (2009). Gender difference in the effect of daytime sleep on declarative memory for pictures. *Journal of Zhejiang University SCIENCE B*, *10*(7), 536–546. https://doi.org/10.1631/jzus.B0820384

Wang, J.-Y., Weber, F. D., Zinke, K., Noack, H., & Born, J. (2017). Effects of sleep on word pair memory in children—Separating item and source memory aspects. *Frontiers in Psychology*, *8*, Article 1533. https://doi.org/10.3389/fpsyg.2017.01533

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

Wilhelm, I., Prehn-Kristensen, A., & Born, J. (2012). Sleep-dependent memory consolidation—What can be learnt from children? *Neuroscience & Biobehavioral Reviews*, *36*(7), 1718–1728. https://doi.org/10.1016/j.neubiorev.2012.03.002

Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, *265*(5172), 676–679. https://doi.org/10.1126/science.8036517

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, *55*, 235–269. https://doi.org/10.1146/annurev.psych.55.090902.141555

Wixted, J. T. (2005). A theory about why we forget what we once knew. *Current Directions in Psychological Science*, *14*(1), 6–9. https://doi.org/10.1111/j.0963-7214.2005.00324.x

Yaroush, R., Sullivan, M. J., & Ekstrand, B. R. (1971). Effect of sleep on memory: II. Differential effect of the first and second half of the night. *Journal of Experimental Psychology*, *88*(3), 361–366. https://doi.org/10.1037/h0030914

Yonelinas, A. P., Ranganath, C., Ekstrom, A. D., & Wiltgen, B. J. (2019). A contextual binding theory of episodic memory: Systems consolidation reconsidered. *Nature Reviews Neuroscience*, *20*, 364–375. https://doi.org/10.1038/s41583-019-0150-4

Acknowledgements

This thesis would not have been possible without the support of many people. First of all, I like to thank Edgar, who provided advice and guidance whenever needed and always encouraged me to go my own way, both inside and outside academia. I also like to thank Bea for her support and the valuable discussions about source memory and forgetting. Further thanks go to Almut, with whom I spent many hours discussing study coding. Moreover, I am thankful to my fellow PhD students, my former colleagues from the GESIS, and the academic and administrative staff from the GESS, the SMiP group, and the chair of Cognitive Psychology and Individual Differences for intellectual and organizational support. I am also grateful to family, friends, and my colleagues at PwC for their emotional support. Many thanks to Emily, who even manages to make my cheeks hurt with laughter during tough times and to Ahsan, who makes every day special. Last but not least, I would like to thank everyone who participated in the studies and shared the early morning and late evening hours in the lab or online with me.

Thank you.

Statement of Originality

1. I hereby declare that the presented doctoral dissertation with the title *The Sleep Benefit in Episodic Memory: Investigating Underlying Mechanisms* is my own work.
2. I did not seek unauthorized assistance of a third party and I have employed no other sources or means except the ones listed. I clearly marked any quotations derived from the works of others.
3. I did not present this doctoral dissertation or parts of it at any other higher education institution in Germany or abroad.
4. I hereby confirm the accuracy of the declaration above.
5. I am aware of the significance of this declaration and the legal consequences in case of untrue or incomplete statements.

I affirm in lieu of oath that the statements above are to the best of my knowledge true and complete.

Signature:

Date:

Co-Authors' Statements

Co-Author: Edgar Erdfelder

I hereby confirm that the following three articles included in the thesis *The Sleep Benefit in Episodic Memory: Investigating Underlying Mechanisms* were primarily conceived and written by Sabrina Berres, PhD candidate at the *Center for Doctoral Studies in Social and Behavioral Sciences (CDSS)* of the *Graduate School of Economic and Social Sciences (GESS)* at the University of Mannheim:

Berres, S., & Erdfelder, E. (2021). The sleep benefit in episodic memory: An integrative review and a meta-analysis. *Psychological Bulletin, 147*(12), 1309–1353. https://doi.org/10.1037/bul0000350.

Berres, S., & Erdfelder, E. (2023). *Is it all about storage? Effects of encoding strength and test expectation on the sleep benefit in episodic memory* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

Berres, S., Erdfelder, E., & Kuhlmann, B. G. (2023). *Does sleep benefit source memory? Investigating 12-hr retention intervals with a multinomial modeling approach* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

I sign this statement to the effect that Sabrina Berres is credited as the primary source of the ideas and the main author of all three articles. She worked out the theoretical background, carried out the study selection and coding, designed and conducted the experiments, performed the analyses, wrote the first drafts, and contributed to the revision of the three manuscripts. I contributed to the conceptualization of the studies and provided methodological advice and supervision throughout the process. I also contributed to revising and improving the manuscripts.

Prof. Dr. Edgar Erdfelder
Mannheim, April 2023

Co-Author: Beatrice G. Kuhlmann

I hereby confirm that the following article included in the thesis *The Sleep Benefit in Episodic Memory: Investigating Underlying Mechanisms* was primarily conceived and written by Sabrina Berres, PhD candidate at the *Center for Doctoral Studies in Social and Behavioral Sciences (CDSS)* of the *Graduate School of Economic and Social Sciences (GESS)* at the University of Mannheim:

Berres, S., Erdfelder, E., & Kuhlmann, B. G. (2023). *Does sleep benefit source memory? Investigating 12-hr retention intervals with a multinomial modeling approach* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

I sign this statement to the effect that Sabrina Berres is credited as the primary source of the ideas and the main author of the article. She worked out the theoretical background, designed and conducted the experiments, performed the analyses, wrote the first drafts, and contributed to the revision of the manuscript. I contributed to the conceptualization of the study and provided methodological advice. I also contributed to revising and improving the manuscript.

Prof. Dr. Beatrice G. Kuhlmann
Mannheim, April 2023

Copies of Articles

The Sleep Benefit in Episodic Memory: An Integrative Review and a Meta-Analysis

Sabrina Berres and Edgar Erdfelder

Department of Psychology, School of Social Sciences, University of Mannheim

Is it all About Storage? Effects of Encoding Strength and Test Expectation on the Sleep Benefit in Episodic Memory

Sabrina Berres and Edgar Erdfelder

Department of Psychology, School of Social Sciences, University of Mannheim

Author Note

Sabrina Berres https://orcid.org/0000-0001-9868-8824

Edgar Erdfelder https://orcid.org/0000-0003-1032-3981

Additional materials (i.e., stimulus material, codebook, data set, analysis code) are provided online on the Open Science Framework (OSF; https://osf.io/xe8rt/?view_only=820b76736cec4762a6ba87d360b8d95d). The two reported experiments were not preregistered. We have no conflicts of interest to disclose. This research was supported by the University of Mannheim's Graduate School of Economic and Social Sciences (GESS) and the Research Training Group "Statistical Modeling in Psychology" (SMiP), funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—GRK 2277. We thank Elena Dülz, Lea Hobrecker, Annalena Knautz, Amelie Maulbetsch, Sophia Peter, Teodora Roydeva, Daniela Wickert, Jana Zettl, and Jahn Zizer for their help in collecting data for Experiment 1.

Correspondence concerning this article should be addressed to Sabrina Berres or Edgar Erdfelder, Department of Psychology, School of Social Sciences, University of Mannheim, A5, Room B 118, 68159 Mannheim, Germany. Email: sberres@mail.uni-mannheim.de or erdfelder@psychologie.uni-mannheim.de

Abstract

People recall more information after sleep than after an equally long period of wakefulness. However, previous research has shown that this sleep benefit is affected by salience cues that tag memories for prioritized consolidation during sleep, suggesting better storage as a consequence of sleep. In two experiments, we evaluated this explanation and explored whether sleep also improves memory retrieval due to reduced retroactive interference during sleep. To disentangle the cognitive processes involved, we used the Encoding-Maintenance-Retrieval (EMR) multinomial model. The model provides separate measures for word-pair association encoding (e), maintenance across retention intervals (m), and free-recall retrieval of stored word pairs (r_f). In Experiment 1, we manipulated salience during encoding via encoding strength, that is, we present some stimuli once and others three times. Our results for memory storage were mixed. Apparently, memory consolidation during sleep improves storage of all memories equally, regardless of their strength. Rather, a larger sleep benefit in memory retrieval for medium compared to strong memories emerged, in line with independent research showing that retroactive-interference effects decrease with encoding strength. In Experiment 2, we varied salience immediately following encoding by manipulating test expectation. Test expectation affected neither storage nor retrieval contributions to the sleep benefit. Importantly, using a 60%-learning criterion, the absence of the sleep benefit in retrieval given strong encoding was replicated. We conclude that sleep improves maintenance in and retrieval from memory. Whereas encoding strength moderates the sleep benefit in retrieval, it does not impact storage benefits significantly.

Word count (abstract): 244 words

Keywords: episodic memory, consolidation during sleep, interference, encoding strength, test expectation

Is it all About Storage? Effects of Encoding Strength and Test Expectation on the Sleep Benefit in Episodic Memory

When learning is followed by sleep compared to an equally long period of wakefulness, people recall more of the learned information. This sleep benefit in episodic memory is a well-established finding in human memory research. It is typically explained by one of two key mechanisms (for a review, see Berres & Erdfelder, 2021). The first mechanism is memory consolidation—a process that converts recently encoded and thus labile memories into more stable long-term memory representations during sleep (Buzsáki, 1998; Diekelmann & Born, 2010; Dudai, 2004, 2012; Dudai et al., 2015; Klinzing et al. 2019; Rasch & Born, 2013). Therefore, memory consolidation actively increases episodic memory storage. The second mechanism is sleep-induced protection from retroactive interference (Ecker, Brown, & Lewandowsky, 2015; Ecker, Tay, & Brown, 2015; Ekstrand, 1967; Jenkins & Dallenbach, 1924; Yonelinas et al., 2019). In contrast to wakefulness, new learning is virtually absent during sleep, thereby reducing retroactive interference for information learned prior to sleep. Because retroactive interference is due to retrieval problems in the first place (e.g., Tulving & Psotka, 1971), the reduction of retroactive interference during sleep passively increases episodic memory retrieval relative to wake control conditions. In fact, recent findings suggest that both mechanisms—better maintenance in memory and better retrieval from memory—might jointly contribute to the sleep benefit (Berres & Erdfelder, 2021; Erdfelder et al., 2022). That is, sleep compared to wakefulness appears to strengthen both episodic memory storage and retrieval.

Despite encouraging past advances towards a better understanding of the sleep benefit, it is still unclear whether and how the underlying cognitive processes are affected by salience cues.[1] It has been argued that salience cues presented during or shortly after encoding may tag memories for preferred memory consolidation during sleep (e.g., Barco et al., 2008; Frey & Morris, 1998; Morris, 2006; Redondo & Morris, 2011). Although the mechanisms by which these tags affect memory

---

[1] We use "salience cues" as an umbrella term for all cues that indicate future stimulus relevance (e.g., emotion, directed forgetting, incentives, encoding strength, test expectation).

consolidation during sleep are not yet well understood, it has been hypothesized that tagging boosts memory consolidation during sleep, resulting in better memory storage (e.g., Antony & Paller, 2017; Bennion et al., 2016; Stickgold & Walker, 2013). In line with this hypothesis, previous research showed larger sleep benefits for (a) emotional versus neutral stimuli (e.g., Hu et al., 2006; Payne et al., 2015), (b) to-be-remembered versus to-be-forgotten information (Rauchs et al., 2011; Saletin et al., 2011), (c) incentivized stimuli (e.g., Oudiette et al., 2013; but see Baran et al., 2013), (d) strongly compared to weakly encoded information (e.g., Schoch et al., 2017; Tucker & Fishbein, 2008; but see Petzka et al., 2021), and (e) stimuli expected to be tested in the future (Wilhelm et al., 2011; but see Ashton & Cairney, 2021).

However, previous research also suggests an additional effect of encoding strength on sleep benefits in subsequent memory retrieval. Specifically, it has been shown that retroactive interference tends to decrease with higher study efforts (Howe, 1970; McGeoch, 1929). This may in turn reduce sleep benefits in retrieval because these benefits rely on relatively strong retroactive interference during wakefulness. Hence, it seems plausible that study efforts manipulated via encoding strength additionally affect sleep-induced reduction of retroactive interference and thus memory retrieval. We therefore investigated sleep benefits not only in episodic memory storage but also in subsequent memory retrieval. In two experiments, we used a multinomial processing tree (MPT) model to disentangle storage and retrieval contributions to memory performance. To further evaluate the consolidation explanation, we selected two salience manipulations—encoding strength (Experiment 1) and test expectation (Experiment 2)—that clearly differ in the timing of salience tagging: While encoding strength (i.e., presentation frequency) is manipulated during learning, test expectation is manipulated immediately following the learning phase.

In sum, the current work aims to advance our understanding of the cognitive processes underlying effects of information salience on the sleep benefit in episodic memory. Whereas the consolidation account of the sleep benefit predicts that information salience boosts sleep benefits in episodic memory storage, it does not allow for clear-cut predictions concerning memory retrieval. By manipulating salience during (encoding strength) and after learning (test expectation), we therefore evaluate the consolidation explanation and explore the role of possible additional retrieval contributions as suggested by research of Howe (1970) and McGeoch (1929).

The Current Experiments

To investigate whether information salience affects (a) sleep benefits in episodic memory storage caused by memory consolidation, (b) sleep benefits in episodic memory retrieval due to protection from retroactive interference, or (c) both, we used the Encoding-Maintenance-Retrieval (EMR) model (Erdfelder et al., 2022; Küpper-Tetzel & Erdfelder, 2012). This model belongs to the class of Multinomial Processing Tree (MPT) models (for reviews, see Batchelder & Riefer, 1999; Erdfelder et al., 2009) and is based on observed frequencies for 12 event categories. These categories result from the combination of all possible outcomes in an immediate cued recall task right after learning, followed by delayed free and cued recall tasks after a longer retention interval (see Appendix A for a more detailed description of the categories).

Based on observed frequencies for these 12 categories, the EMR model provides separate measures for successful encoding of word-pair associations (e), maintaining encoded associations across the retention interval (m), and retrieving stored associations in free recall (r_f). Further parameters are estimated for the probability of word-pair retrieval in cued recall (r_c) and various probabilities of single word retrieval during free recall when word-pair associations are stored (s_s and s_u) versus not stored in memory (u_s and u_u). The index “s” versus “u” of the latter parameters indicates whether the preceding cued recall attempt was successful versus unsuccessful, respectively. Please refer to Appendix A for a more detailed description of the model.

Following Erdfelder et al.’s (2022) recommendation, we first fitted a parsimonious independence model which precludes effects of the preceding memory test and includes 6 parameters only (i.e., parameter e, m, r_c, r_f, s, u). In case this independence model produced misfit, we relaxed the independence assumptions in a second step, starting with parameter u followed by parameter s (cf. Erdfelder et al., 2022). Importantly, because we also wanted to check for potential differences in the probability of retrieving stored associations in cued recall (i.e., parameter r_c), we did not apply the recommended restriction $r_c = r_{cs} = r_{cu}$ (see Appendix A) in our model-based analyses.

As such, the EMR model has been successfully applied in sleep-related and non-sleep-related research (e.g., Erdfelder et al., 2022; Küpper-Tetzel & Erdfelder, 2012; Küpper-Tetzel et al., 2014; Quevedo Pütter & Erdfelder, 2022). Moreover, the model's psychological validity was demonstrated by showing that the model parameters can be selectively influenced by experimental manipulations known to affect specific cognitive processes. For example, Erdfelder et al. (2022) showed that the length of the retention interval affects storage (m) and free recall retrieval (r_f), but not encoding parameters (e). Furthermore, providing category names as retrieval cues during free recall selectively influenced the parameter for free recall retrieval (r_f) while leaving the parameters for encoding (e) and maintenance (m) unaffected (Erdfelder et al., 2022, Experiment 2). Thus, there is evidence that the core parameters of the EMR model validly measure the memory processes they are supposed to measure.

MPT models are most often fitted to empirical data using complete or partial pooling approaches (for a tutorial, see O. Schmidt et al., 2023). In the complete pooling approach, observed category frequencies are aggregated across participants and typically analyzed using maximum likelihood (ML) parameter estimation. By contrast, partial pooling typically employs a Bayesian approach that relies on Markov-chain Monte Carlo (MCMC) methods (Heck et al., 2018) to estimate individual and group-level MPT-parameters, thereby explicitly accounting for potential individual differences between participants. Following Erdfelder et al. (2022), we used both approaches to check whether our results are robust against the different distributional assumptions involved in complete and partial pooling. We made use of the software multiTree (Moshagen, 2010) and the latent-trait model (Klauer, 2010) as implemented in the R package TreeBUGS (Heck et al., 2018) for complete and partial pooling approaches, respectively.

For the complete pooling approach conducted with multiTree, we allowed for up to 10,000 iterations for the expectation maximization (EM) parameter estimation algorithm. Model fit was assessed using the goodness of fit statistic G^2. To evaluate differences between experimental conditions, we used the difference in the goodness of fit statistic between baseline and alternative models (ΔG^2). For the partial pooling approach conducted with TreeBUGS, we used 200,000 iterations from four MCMC chains, keeping every 10th iteration. Moreover, we used 100,000 iterations for adaptation to adjust MCMC sampling in the R package JAGS (Plummer, 2003) and removed 100,000 burn-in iterations. Furthermore, we used weakly informative priors (i.e., the default priors in TreeBUGS, see Heck et al., 2018). Stimulus materials, codebooks, data sets, and analysis codes for both experiments reported here are available on the OSF (https://osf.io/xe8rt/?view_only=820b76736cec4762a6ba87d360b8d95d). The two experiments were not preregistered.

Experiment 1

Everyday life comes with a multitude of information that cannot be encoded equally well due to limited cognitive resources. Consequently, some pieces of information are better encoded and thus possess higher memory strength than others. For this reason, researchers investigated whether using encoding strength as a salience cue affects how much episodic memories profit from sleep. According to the consolidation account, the sleep benefit in episodic memory storage follows an inverted U-shaped function of memory strength, with the largest sleep benefit for medium degrees of encoding strength (Stickgold, 2009). In line with this prediction, there is growing empirical support that a certain minimum level of memory strength at encoding is necessary for the sleep benefit to occur (e.g., Denis et al., 2020; Muehlroth et al., 2020; Rauchs et al., 2011). That is, floor effects might counteract sleep-induced memory consolidation when encoding strength is very low (Petzka et al., 2021). Also in line with the prediction, sleep benefits in episodic memory appear to decrease when encoding strength increases beyond a medium level up to very high levels. For example, electroencephalography (EEG) studies showed that memories of medium strength (i.e., approximately 50% mean correct responses in immediate recall tasks; see Table 1) are associated with higher spindle density during non-rapid eye movement (NREM) sleep (C. Schmidt et al., 2006, but see Denis et al., 2021), which is thought to promote memory consolidation (e.g., Fernandez & Lüthi, 2020; Klinzing et al., 2019; Muehlroth et al., 2020; Peyrache & Seibt, 2020). In addition, targeted memory reactivation (TMR) studies showed that TMR during sleep yielded better memory performance provided that initial recall performance is moderately (but not highly) accurate (Cairney et al., 2016; Creery et al., 2015).[2]

Further support for an inverted U-shaped curve of the sleep benefit in episodic memory storage comes from behavioral results. Note, however, that the terms “weak encoding” and “strong encoding” have typically been used relative to each other in a quite vague sense. Also, various encoding-strength manipulations have been employed. To illustrate, researchers presented study materials once versus multiple times (Denis et al., 2020; Denis et al., 2021), used immediate tests versus no tests with feedback versus no feedback after learning (Schoch et al., 2017; Tucker & Fishbein, 2008; Ukraintseva & Dorokhov, 2012), or applied different learning criteria (Sheth et al., 2012) and presentation times for feedback (Drosopoulos et al., 2007). For this reason, “weak encoding” in one experiment may correspond to “strong encoding” in another experiment or vice versa. To enable comparison of previous research with the results of the current experiments, we refer to the percentage of correct responses in immediate recall, as comprehensively summarized in Table 1. We refer to 40%-60% correct immediate recall responses as medium encoding strength. According to this criterion, studies with $\lesssim$ 50% mean correct immediate recall correspond to the lower limb and studies with $\gtrsim$ 50% to the upper limb of Stickgold’s (2009) inverted U-shaped function. Applying this categorization scheme to the studies, the overall result pattern supports the consolidation account: Most studies assigned to the lower limb showed a smaller sleep benefit in weak compared to stronger encoding conditions (e.g., Schoch et al., 2017; Tucker & Fishbein, 2008; but see Denis et al., 2020; Denis et al., 2021; Ukraintseva & Dorokhov, 2012). By contrast, most

---

[2] In targeted memory reactivation (TMR) studies, subtle auditory reminder cues are presented during sleep. The auditory cues are presented jointly with the learning material during the study phase and are played again during sleep to reactivate the associated memories. This reactivation should lead to better memory performance in a delayed memory test.

studies falling in the upper limb of the inverted U-shaped function showed a larger sleep benefit in medium- compared to strong-encoding conditions (e.g., Cairney et al., 2016; Drosopoulos et al., 2007; Petzka et al., 2021; Sheth et al., 2012; but see Denis et al., 2020; Denis et al., 2021; for a summary, see Table 1). Overall, the empirical evidence from EEG, TMR, and behavioral studies are in line with the consolidation account. According to this account, very strong memories should benefit less from further strengthening through memory consolidation during sleep, most likely because they already possess sufficient memory strength by default (Denis et al., 2020; Denis et al., 2021; Petzka et al., 2021).

Despite the plausibility of the consolidation account, there is also recent behavioral evidence suggesting that memory consolidation during sleep benefits all memories irrespective of their memory strength (Petzka et al., 2021). Specifically, Petzka et al. (2021) examined the sleep benefit for medium and strong memories (i.e., approximately 65% vs. 93% mean correct responses in immediate recall tasks; see Table 1) in two conditions. The first condition was a relatively easy standard test condition with a delayed memory test following the retention interval. The second condition—the difficult test condition—additionally introduced retroactive interference before the delayed memory test. Whereas a sleep benefit for medium but not for strong memories emerged in the standard test condition, all memories benefitted from sleep compared to wakefulness in the difficult test condition. As a possible interpretation, the authors suggest that sleep may foster episodic memory storage irrespective of memory strength unless concealed by ceiling effects for strong memories like in the relatively easy standard-test condition (Petzka et al., 2021). In the current experiment, we test Petzka et al.’s (2021) interpretation thoroughly by disentangling storage and retrieval contributions to the sleep benefit with the EMR model.

Even though the empirical evidence that has accumulated so far supports the consolidation explanation, this does not preclude that encoding strength may additionally affect sleep-induced reduction of retroactive interference and thus memory retrieval. In fact, Erdfelder et al. (2022) showed that sleep compared to wakefulness has two qualitatively different effects on episodic memory: It actively boosts memory storage due to memory consolidation and, in addition, improves memory retrieval passively by reducing retroactive interference. Also, as already pointed out above, previous studies showed that retroactive interference tends to decrease with higher study efforts (Howe, 1970; McGeoch, 1929), suggesting that sleep benefits in episodic memory retrieval may be reduced for more strongly encoded memories. In Experiment 1 we therefore investigated whether encoding strength manipulated in the medium to high range affects (a) sleep benefits in episodic memory storage, (b) sleep benefits in episodic memory retrieval, or (c) both. In line with Stickgold (2009), we expected larger sleep benefits in episodic memory storage for the medium- compared to the strong-encoding condition.

Method

We applied a 2 × 2 mixed factorial design with encoding strength (medium vs. strong encoding) as within-subject factor and wake versus sleep as between-subjects factor.

Participants

Data collection took place in spring 2018. We aimed for a sample size of at least 24 participants (i.e., $n_{\text{sleep}} = 12$, $n_{\text{wake}} = 12$) but tried to collect as many participants as possible before the semester break. The minimum target sample size was based on previous studies investigating encoding-strength effects on the sleep benefit (see Table 1). For example, Drosopoulos et al. (2007) and Sheth et al. (2012) used 10 participants per group to demonstrate a larger sleep benefit in medium-compared to strong encoding conditions.

We performed a priori sensitivity analyses with (a) G*Power (Faul et al., 2007) for the ANOVA-based memory performance analysis, and (b) multiTree for the EMR model-based analysis. All sensitivity analyses were performed using a conventional α-level of .05 and the minimum target sample size of $N = 24$ participants (i.e., $n_{\text{sleep}} = 12$, $n_{\text{wake}} = 12$). For the ANOVA-based memory performance analysis, sensitivity analyses confirmed that ANOVA $F(1, 22)$ tests can detect effects exceeding medium strength (i.e., $f = .30$) with a power of $1 - \beta = .80$ (Faul et al., 2007). For the EMR model-based analysis, we performed a sensitivity analysis with 40 word pairs per participant and encoding-strength condition (i.e., 2 x 480 = 960 observations in total). Because this is the first time the EMR model is used to investigate encoding-strength effects on the sleep benefit, we assumed a minimum difference of 0.10 between the wake and sleep group for m- and r_f-parameters in medium- and strong-encoding conditions, respectively. This assumption was based on the results by Erdfelder et al. (2022) who applied the EMR model for the first time to investigate sleep benefits in episodic memory. The authors found sleep benefits of ≥ 0.10 in m- and r_f-parameters for strong encoded memories (i.e., > 72% mean correct responses in immediate cued recall tasks). The sensitivity analysis resulted in a power (1 - β) of .68 for parameter m and .94 for parameter $r_{f,}$ respectively.[3] As baseline parameter values for the sensitivity analyses, we used the ML parameter estimates of Erdfelder et al.'s (2022) no-cue wake condition in Experiment 2 (e = .75, m = .85, r_c = .98, r_f = .51, s = .09, u_s = .25, u_u = .07).

In total, 69 persons participated in the laboratory experiment at the University of Mannheim. They were recruited via campus advertisements, social media, or personal contacts and received home-made snacks for their participation. We excluded two participants because they reported active rehearsal of word pairs during the retention interval. Furthermore, we applied the exclusion criteria recommended for the EMR model (cf. Erdfelder et al., 2022) and excluded 14 participants with less than 35% and 26 participants with more than 95% correct responses in immediate cued recall (see Appendix A for a detailed explanation). Thus, for data analysis 27 participants remained, 12 in the sleep and 15 in the wake group. The participants were between 17 and 51 years of age (*M* = 22.89 years, *SD* = 6.46), 20 (74.07%) were female.[4] All were native or fluent speakers of German and, except for one person, university students.

Procedure

---

[3] Note that assuming a difference of 0.10 between the wake and sleep group in e- and r_c-parameters resulted in a power (1 - β) of .95 for parameter e and .99 for parameter r_c, respectively.

[4] The 17-year-old person showed a signed parental permission to participate upon entering the laboratory. Only one participant was more than 35 years old. Importantly, similar results emerged for memory performance and model-based analyses when including only participants between 18 and 35 years of age (see Tables S2 to S5 in the supplemental materials).

The random assignment took place during online registration with SoSci Survey (Leiner, 2018). Participants randomly assigned to the wake condition were informed that the experiment consists of two sessions with the first session at 8:30 a.m., followed by the second session 12 hr later. For the sleep condition, starting times of the two sessions were reversed. In both conditions, participants were asked to choose a date for participation.

Both sessions of the experiment were conducted with OpenSesame (Mathôt et al., 2012) in the laboratory. In the first session, participants read and agreed to the information consent before they started the experiment answering several questions. Specifically, participants were asked whether they slept at least 6 hr in the night before the experiment and when they woke up this morning. Participants also indicated on the Stanford Sleepiness Scale (SSS, Hoddes et al., 1973) how sleepy they feel right now. After completing the questions, participants learned 80 weakly related word pairs with the instruction to memorize all word pairs as well as possible for a cued recall task at the end of the session. To manipulate encoding strength, half of the word pairs were presented once, the other half three times in random order. Note that we ensured not to present word pairs multiple times in a row, that is, at least one other word was presented between repetitions of the same word. To ensure at least medium memory strength even when word pairs are presented only once, each word pair was presented for 5 s with an interstimulus interval of 100 ms. The learning phase was followed by a distractor task of 15 math equations (e.g., 24 / 2 + 32 = 44). By pressing one of two keys, participants decided whether an equation was right or wrong. When responding within 8 s, participants received feedback on the correctness of their answer. Otherwise, they were informed that they responded too slowly. After the distractor task, participants performed the previously announced immediate cued recall task. In this self-paced task, the left word of a word pair was presented, and participants had to remember and type in the right word (or leave the field blank in case of recollection failure). Finally, we informed the participants that in the next session they will receive the same cued recall task again, in addition to other tasks. We also asked them (a) not to actively rehearse learned word pairs, (b) not to consume alcoholic drinks, (c) not to take a nap during the day, and (d) to go to bed at their usual time. During the 12-hr retention interval, participants pursued their usual daily activities or slept at home. To obtain a rough measure of total sleep duration and the time spent in deep and light sleep, participants in the sleep condition wore a commercial fitness tracker (Xiaomi Mi Band 2).

At the beginning of the second session, the SSS was presented and answered by the participants. This was followed by a surprise free recall task. In this paper-pencil task, participants were asked to write down all learned word pairs they could remember within 10 min. In case they remember only one word of a pair, participants were instructed to write it down as well. Moreover, we notified participants that the order in which word pairs or the two words of a pair are written down does not matter. After reading the instructions, a countdown was started on the computer screen and participants filled in the sheet of paper with two columns and 80 numbered rows. The free recall task was followed by a delayed cued recall task. This task was identical to the immediate cued recall except for the presentation order of the left word, which was randomized. Thereafter, participants filled in the German version of the reduced Morningness-Eveningness Questionnaire (rMEQ; Randler, 2013) and answered demographic questions. We also asked participants in the sleep condition about last night's sleep quality and the time they fell asleep and woke up. By contrast, participants in the wake condition were asked whether they slept during the last 12 hr, and if so, when they started and stopped sleeping. In addition, we asked participants in both conditions whether and how much alcoholic drinks they consumed, and whether they actively rehearsed word pairs during the retention interval. The testing session ended with a question about their belief concerning the study aim. Finally, participants were thanked, debriefed, and compensated.

Material

To facilitate learning and to ensure that responses in recall tasks are based on memory retrieval rather than guessing, we used weakly associated word pairs as study material. All words were concrete German nouns taken from three category-production norms (Flammer et al., 1985; Mannhaupt, 1983; Scheithe & Bäuml, 1995). We created 80 word pairs by selecting an untypical word as cue (e.g., candy) and a typical word as target (e.g., bread) from 80 categories (e.g., food). Whenever possible, we selected a cue word with a production index below .02 and a target word from the four most frequently named words of a category. The word pairs were than randomly assigned to either List A or List B, so that both lists contained 40 word pairs each. We counterbalanced the two lists across participants in wake and sleep conditions to minimize material effects. Specifically, half of the participants in both experimental conditions learned List A once and List B three times whereas the other half learned List B once and List A three times. All word pairs including production indices and the assigned list are available on the OSF (https://osf.io/xe8rt/?view_only=820b76736cec4762a6ba87d360b8d95d).

Results

We applied a significance level of α = .05 for all analyses. The sample characteristics are presented in Table S1 in the supplemental materials. All participants stated that they did not consume any alcoholic drinks between the two sessions of the experiment. Importantly, all participants reported sleeping for at least 6 hr in the night before the experiment. Despite the request not to sleep during the day, two participants in the wake condition took a nap.[5] For the night between learning and testing sessions, participants in the sleep condition reported a mean sleep duration of 7.54 hr (*SD* = 0.61). This corresponds to the sleep duration assessed by fitness trackers (sleep duration: *M* = 7.84 hr, *SD* = 0.67; time awake: *M* = 0.01 hr, *SD* = 0.02; deep sleep: *M* = 2.04 hr, *SD* = 1.15; light sleep: *M* = 5.79 hr, *SD* = 1.29). In addition, participants in the sleep condition reported medium sleep quality on a scale ranging from 1 (*very bad*) to 5 (*very good*; *M* = 3.17, *SD* = 0.39).

Memory Performance Analysis

To check whether the encoding-strength manipulation was successful, we conducted a mixed ANOVA on cued recall immediately following learning, with medium- versus strong-encoding

---

[5] One participant napped for 28 min, the other for 1 hr. Note that similar results emerged for memory performance and model-based analyses when excluding the two participants from the wake condition (see Tables S6 to S9 in the supplemental materials).

strength as within-subject factor and wake versus sleep as between-subjects factor. Sample means and standard errors for the wake and sleep groups and the two encoding-strength conditions are shown in Table 2. As expected, the main effect of encoding strength was clearly statistically significant, $F(1, 25) = 111.33$, $MSE = 17.01$, $p < .001$, $\eta_p^2 = .82$, with better immediate cued recall performance for strong compared to medium encoding. Also as expected, neither the main effect of wake versus sleep, $F(1, 25) = 0.00$, $MSE = 47.56$, $p = .993$, $\eta_p^2 < .001$, nor the interaction of encoding strength and wake versus sleep, $F(1, 25) = 0.27$, $MSE = 17.01$, $p = .610$, $\eta_p^2 = .01$, were statistically significant. Most importantly, the mean immediate recall rate of 56.58% and 86.38% for word pairs presented once versus three-times, respectively, confirmed the effectiveness of our encoding strength manipulation (see Table 1). Together these results indicate that medium- versus strong encoding was manipulated successfully.

To test whether the sleep benefit is larger in the medium- compared to the strong-encoding condition as predicted for the upper limb of the inverted U-shaped function, we again used mixed ANOVAs, this time focusing on delayed cued recall, cued recall forgetting (i.e., the difference initial - delayed cued recall performance), and free recall scores as dependent variables. For delayed cued and free recall, we found a statistically significant main effect of encoding strength, delayed cued recall: $F(1, 25) = 157.28$, $MSE = 12.24$, $p < .001$, $\eta_p^2 = .86$; free recall: $F(1, 25) = 60.48$, $MSE = 13.07$, $p < .001$, $\eta_p^2 = .71$. In both recall tasks, memory performance was better for strong- compared to medium encoding (see Table 2). We observed no statistically significant main effect of wake versus sleep, delayed cued recall: $F(1, 25) = 0.69$, $MSE = 65.08$, $p = .414$, $\eta_p^2 = .03$; free recall: $F(1, 25) = 0.57$, $MSE = 36.82$, $p = .459$, $\eta_p^2 = .02$. The interaction of encoding strength and wake versus sleep was also not statistically significant, delayed cued recall: $F(1, 25) = 0.73$, $MSE = 12.24$, $p = .402$, $\eta_p^2 = .03$; free recall: $F(1, 25) = 1.09$, $MSE = 13.07$, $p = .307$, $\eta_p^2 = .04$. Although the sleep benefit in free recall appears to be descriptively larger for medium compared to strong encoding, this interaction effect was not statistically significant.

For cued recall forgetting, no statistically significant main effect of encoding strength emerged, $F(1, 25) = 0.05$, $MSE = 2.89$, $p = .832$, $\eta_p^2 = .002$. However, the main effect of wake versus sleep was statistically significant, $F(1, 25) = 8.11$, $MSE = 5.63$, $p = .009$, $\eta_p^2 = .25$, showing less forgetting in the sleep than in the wake condition (see Table 2). By contrast, the interaction of encoding strength and wake versus sleep was not statistically significant, $F(1, 25) = 0.25$, $MSE = 2.89$, $p = .621$, $\eta_p^2 = .01$. In the overall memory performance measures, we thus found no significant evidence for a larger sleep benefit in medium compared to strong encoding, neither in delayed cued recall, nor in cued recall forgetting or in free recall.

Due to the application of the strict exclusion criteria recommended for the EMR model, 40 participants unfortunately had to be excluded. To assess the possible impact of this exclusion on the memory performance data, we performed all ANOVAs for the memory performance analysis again, this time including the 40 previously excluded participants (i.e., new sample size: $n_{\text{sleep}} = 35$, $n_{\text{wake}} = 32$). Importantly, similar results emerged when using the larger sample size of $N = 67$ (for details, see the supplemental materials).

Model-Based Analysis

As outlined in The Current Experiments section, we first fitted the independence variant of the EMR model which, however, produced misfit, $G^2(20) = 40.34$, $p = .005$. In a second step, we therefore relaxed the independence assumption for parameter u ($u = u_s = u_u$). The resulting 7-parameter EMR model version fitted the aggregated data well, $G^2(16) = 21.31$, $p = .167$. Table 3 summarizes the ML parameter estimates and standard errors of the 7-parameter EMR model for the four experimental conditions.

First, we checked whether the encoding-strength manipulation was successful. As expected, the estimated probability of successfully encoding word-pair associations (e) was approximately 30% larger in strong-encoding than medium-encoding conditions, indicating a statistically significant main effect of encoding strength, $\Delta G^2(1) = 435.35$, $p < .001$. There were neither statistically significant encoding differences between wake and sleep conditions, $\Delta G^2(1) = 0.02$, $p = .885$, nor a statistically significant interaction of encoding strength and wake versus sleep, $\Delta G^2(1) = 0.46$, $p = .499$.[6] Thus, in line with the memory performance analysis, these results show that encoding strength was successfully manipulated.

Second, we tested the hypothesis whether the sleep benefit in episodic memory storage is larger for medium-encoded memories compared to strong-encoded memories. For storage parameter m, we found a statistically significant main effect of encoding strength, $\Delta G^2(1) = 4.12$, $p = .042$, indicating an approximately 3% larger maintenance estimate for strong- than medium-encoding conditions. The main effect of wake versus sleep was also statistically significant, $\Delta G^2(1) = 43.03$, $p < .001$, showing an increase of about 7% in the estimated probability to successfully maintain encoded associations across the retention interval in the sleep relative to the wake condition. By contrast, we found no statistically significant interaction of encoding strength and wake versus sleep, $\Delta G^2(1) = 0.58$, $p = .445$ (for the test method, see Footnote 3). Although this result pattern is descriptively in line with the consolidation account, the alternative hypothesis that sleep improves storage for all memories equally, irrespective of their strength, cannot be ruled out.

Third, because protection from retroactive interference might also contribute to encoding-strength effects on the sleep benefit, we investigated whether encoding strength and sleep versus wakefulness affect the probability of successfully retrieving stored associations in free recall (r_f). As such, free recall tasks are best suited to capture retrieval processes because they rely more on memory retrieval than cued recall or recognition tasks (cf. Dyne et al., 1990; McKinney, 1935; Postman, 1952). A statistically significant main effect of encoding strength was observed, with about 10% larger estimated retrieval probabilities for strong compared to medium encoding, $\Delta G^2(1) = 19.95$, $p < .001$. Moreover, we found a statistically significant main effect of wake versus sleep, $\Delta G^2(1) = 224.42$, $p < .001$, and a statistically significant interaction of encoding strength and wake versus sleep, $\Delta G^2(1) = 5.52$, $p = .019$. Specifically, whereas retrieval estimates were 9% larger after

---

[6] Interaction tests for MPT models were performed using the method outlined by Kuhlmann et al. (2019).

sleep compared to wakefulness for medium encoding, no sleep benefit emerged for strong encoding because of better retrieval in the wake condition.

As expected, there were no significant differences in cued recall retrieval (r_c), $\Delta G^2(3) = 1.19$, $p = .756$, with participants retrieving stored associations with a probability of about .98 irrespective of condition. Our findings thus replicate analogous results of Erdfelder et al. (2022), showing that cued recall retrieval is indeed often successful so that r_c-parameters are close to 1.

We also conducted the model-based analysis with Klauer's (2010) latent-trait approach as implemented in TreeBUGS (Heck et al., 2018) to check the robustness of our results. As is apparent in Table B1 of the Appendix, the estimated group-level means closely resemble those reported in Table 3, indicating that the result pattern is robust against complete versus partial pooling approaches.

Discussion

Overall, the results of the memory performance analysis did not support the prediction of a larger sleep benefit for medium compared to stronger encoded memories. Specifically, there was no significant interaction effect of encoding strength and sleep versus wakefulness, neither in delayed cued recall, nor in cued recall forgetting or free recall. However, free recall performance showed a descriptive result pattern in the predicted direction. Notably, the lack of a significant interaction in cued recall forgetting replicates Denis et al.'s (2020) findings in 12-hr conditions for successfully visualized items.[7] While these authors found less forgetting after sleep than after wakefulness, this sleep benefit did not interact with presentation frequency (i.e., 2- vs. 4-times-presentation condition). In line with this, our results for pre-post differences in cued recall also showed an almost equally strong sleep benefit for medium and strong encoding.

In contrast to cued recall tasks, we found a descriptive but not statistically significant larger sleep benefit for medium compared to strong encoding in free recall. Because free recall tasks depend

---

[7] As already outline above, Denis et al. (2020) asked participants to visualize the stimuli which were presented for 1.50 s only. Because successfully visualized compared to not visualized stimuli possess higher memory strength at encoding, the former are more suited for a comparison with our results than the latter.

more on memory retrieval as memory cues are lacking, potential effects of encoding strength on sleep-induced protection from retroactive interference likely affect free recall more than cued recall. This might explain why the descriptive trend of a larger sleep benefit for medium compared to strong encoding is limited to free recall.

For a more thorough test of the consolidation account predictions concerning the upper limb of the inverted U-shaped function and for a closer look at the processes underlying encoding-strength effects on the sleep benefit, we now turn to the model-based results. First, we found no significant interaction of encoding strength and wake versus sleep for memory storage. Because we based the minimum target sample size on previous studies, it might be that an interaction effect is present nevertheless, but smaller than results of prior research suggest—particularly because these studies did not use MPT modeling to disentangle storage and retrieval contributions. Of course, the insignificant interaction is also in line with the alternative view that memory consolidation during sleep benefits all memories equally irrespective of their memory strength. This hypothesis was suggested by Petzka et al. (2021) who examined the sleep benefit for medium versus strong encoded memories in a relatively easy standard-test condition and in a difficult-test condition. The authors found a sleep benefit for medium but not for strong encoded memories in the standard-test condition, whereas all memories benefitted from sleep in the difficult-test condition. However, in contrast to Petzka et al.'s (2021) study, the mean success rate in immediate cued recall for strong encoding is comparably low in our current experiment (i.e., approximately 93% vs. 86%; see Table 1). Consequently, it is unlikely that a ceiling effect occurred in the current experiment. To summarize, if memory consolidation during sleep indeed preferably improves memory storage for medium- compared to strong-encoded memories, this effect appears to be quite small and in general hard to detect.

Second, in addition to evaluating the consolidation account, we also explored whether sleep benefits in memory retrieval are affected by encoding strength. We found that sleep compared to wakefulness improves not only episodic memory storage but also episodic memory retrieval. This is in line with recent findings suggesting that storage and retrieval might jointly contribute to the sleep benefit (Berres & Erdfelder, 2021; Erdfelder et al., 2022). More importantly, we found that the sleep benefit in episodic memory retrieval diminishes with increasing encoding strength as might be expected based on previous research on retroactive interference (cf. Howe, 1970; McGeoch, 1929). Apparently, encoding strength affects sleep benefits in episodic memory retrieval more than in episodic memory storage. As such, this finding has important implications for the interpretation of sleep benefits observed in memory performance analyses. Typically, the results of these analyses have been interpreted in terms of the consolidation explanation, thereby neglecting possible effects of sleep on memory retrieval altogether.

In sum, by manipulating encoding strength, we found mixed evidence for memory storage predictions of the consolidation account and demonstrated for the first time that sleep benefits in retrieval from episodic memory are affected by encoding strength. However, as sketched above, tagging of memories for preferred memory consolidation during sleep can take place not only during but also shortly after encoding (e.g., Barco et al., 2008; Frey & Morris, 1998; Morris, 2006; Redondo & Morris, 2011). We therefore manipulated test expectation after learning in a second experiment to further evaluate the consolidation explanation. Moreover, because Experiment 1 appears to be the first demonstration that the sleep benefit in episodic memory retrieval may vanish if encoding strength is high, we also checked in Experiment 2 whether this finding can be conceptually replicated by applying a 60%-learning criterion in immediate cued recall.

Experiment 2

According to the assumption that relevant information is tagged for preferred memory consolidation not only during but also shortly after encoding (e.g., Barco et al., 2008; Frey & Morris, 1998; Morris, 2006; Redondo & Morris, 2011), the sleep benefit in episodic memory storage should be larger in expected than in unexpected delayed memory tests. Specifically, in Wilhelm et al.'s (2011) first experiment, a sleep benefit in cued recall performance occurred only when participants expected a delayed memory test, but not when the delayed test was unexpected. In their study, the authors asked young adults to learn 40 semantically related word pairs. Each word pair was presented for 5 s. Following word-pair presentation, participants repeated a cued recall task until they mastered at least 60% of the pairs. In this immediate cued recall task, each response was followed by the presentation of the correct target word for 2 s. Test expectation was manipulated between participants following word-pair encoding at the end of the learning session. Thus, the test-expectation manipulation could not affect encoding strength. The authors manipulated test expectation by correctly informing one group of participants that memory for the word pairs learned previously will be tested after the retention interval (expectation condition). In contrast, participants in the no-expectation condition were misleadingly informed that they will perform a different task in the testing session. The learning session was followed by a 9-hr retention interval for which sleep versus wakefulness was manipulated between participants. Finally, in the testing session, participants performed a delayed cued recall task and indicated whether they expected a delayed memory test. Wilhelm et al. (2011) showed that sleep compared to wakefulness improves delayed cued recall performance when the memory test is expected but not when it is unexpected. The authors explained their findings with enhanced memory consolidation during sleep for memories that were tagged to be of future relevance. To test this interpretation, we investigated whether the sleep benefit in episodic memory storage is larger when the delayed memory test is expected rather than unexpected.

A further purpose of Experiment 2 was to examine whether the sleep benefit decrease in the strong-encoding condition of Experiment 1 can be replicated when applying a 60%-learning criterion. As such, a criterion of 60% has often been used to ensure sufficient learning and memory strength in sleep and memory research (e.g., Backhaus & Junghanns, 2006; Feld et al., 2013; Fenn & Hambrick, 2012, 2013; Gais & Born, 2004; Marshall et al., 2004; Plihal & Born, 1997; Wilhelm et al., 2011). Hence, to mimic the strong-encoding condition in Experiment 1 and to follow Wilhelm et al. (2011), we implemented a 60%-cutoff for the immediate cued recall task.

In sum, in Experiment 2 we investigated (a) whether the sleep benefit in episodic memory storage is affected by test expectation and (b) whether the findings in the strong-encoding condition of Experiment 1 can be conceptually replicated. Following Wilhelm et al. (2011), we expected a larger sleep benefit in memory storage when the upcoming memory test is expected than when it is not expected.

Method

We used a 2 × 2 factorial design with test expectation (no expectation vs. expectation) and wake versus sleep as between-subjects factors.

Participants

Following the sensitivity analyses performed for the EMR model-based analysis in Experiment 1, we strove for a sample of at least 12 participants per experimental condition, that is, 48 participants in total. Using this minimum target sample size of $N = 48$ participants, an a priori sensitivity analysis with a conventional α-level of .05 for the ANOVA-based memory performance analysis confirmed that ANOVA $F(1, 44)$ tests can detect large effects (i.e., $f = .40$) with a power of $1 - \beta = .80$ (Faul et al., 2007). Hence, the minimum target sample size would be sufficient to find test-expectation effects of the size reported by Wilhelm et al. (2011; i.e. $d = 0.86$ or, equivalently, $f = .43$) for the sleep benefit in episodic memory.

The laboratory experiment took place at the University of Mannheim in spring 2018. As in Experiment 1, we recruited as many participants as possible before the semester break via advertisements on campus and social media. Participants received course credit or money for their participation (i.e., a flat rate of 14.00 €). From 91 participants, we excluded seven participants because they actively rehearsed word pairs during the retention interval. In addition, we excluded four participants with more than 95% correct responses in immediate cued recall by following the recommended exclusion criteria for the EMR model (cf. Erdfelder et al., 2022; see Appendix A for a detailed explanation). Moreover, nine participants of the no-expectation condition had to be excluded because they reported that they expected a delayed memory test. Finally, we excluded one participant of the expectation condition who reported that the delayed memory test was unexpected. Thus, 70 participants remained for data analysis. Participants were between 18 and 54 years of age (*M* = 22.06 years, *SD* = 4.81), 51 (72.86%) were female.[8] They were all native or fluent speakers of German and, except for four persons, university students.

Procedure

The procedure followed that of Experiment 1 and used the same study builders (i.e., SoSci Survey, OpenSesame) with some exceptions: First, after random assignment to one of four experimental groups, participants in the wake condition could choose between 8:00 a.m. and 8:30 a.m., whereas participants in the sleep condition could choose between 8:00 p.m. and 8:30 p.m. for the starting time of the first session. Second, participants learned 40 weakly related word pairs, presented in random order for 5 s each with an interstimulus interval of 100 ms, following the same learning instruction as in Experiment 1. Third, the sequence of stimulus presentation, distractor task, and immediate cued recall task was repeated until participants mastered at least 60% of the word pairs in immediate cued recall. Fourth, test expectation was manipulated between participants at the end of the learning session. Specifically, we informed participants that the experiment aims at assessing motor and verbal learning in the morning and evening. In no-expectation conditions, participants were told that we randomly assigned them to two groups: Group A, which completes the motor-learning session in the morning and the verbal-learning session in the evening, and Group B for which the timing of motor- and verbal-learning sessions was reversed. We informed participants in the no-expectation-wake condition that they were assigned to Group B and participants in the no-expectation-sleep condition that they were assigned to Group A. Both groups were notified that the next session consists of a motor learning task. In test-expectation conditions, participants were told that they will perform the same cued recall task in addition to a motor learning task in the next session. Finally, we asked only participants in expectation conditions not to actively rehearse the

---

[8] Note that only one participant was more than 35 years old. However, comparable results emerged for memory performance and model-based analyses when including only participants between 18 and 35 years of age (see Tables S13 to S16 in the supplemental materials).

learned word pairs during the retention interval as they expected a memory test after the retention interval. In the second session, all participants performed a free and a cued recall task as in Experiment 1, followed by a mock motor task.

Material

Using the procedure of Experiment 1, 40 new weakly associated word pairs were created. Specifically, we picked 40 of the 80 categories used in Experiment 1 and selected a new untypical word as cue and a new typical word as target for each category. All words were concrete German nouns taken from three category-production norms (Eckes, 1985/1994; Flammer et al., 1985; Mannhaupt, 1983). All word pairs including production indices are available on the OSF (https://osf.io/xe8rt/?view_only=820b76736cec4762a6ba87d360b8d95d).

Results

Again, we used α = .05 for all analyses. The sample characteristics are presented in Table S12 in the supplemental materials. All participants except one stated that they did not consume any alcoholic drinks between the two sessions of the experiment. Because the consumed amount of ca. 2.4% (vol/vol) alcohol in 0.33 L was not substantial, we included the person in the analysis. Moreover, except for nine participants ($n_{\text{no expectation, wake}} = 2$, $n_{\text{no expectation, sleep}} = 3$, $n_{\text{expectation, wake}} = 3$, $n_{\text{expectation, sleep}} = 1$), all participants reported sleeping for at least 6 hr in the night before the experiment. Despite the request not to sleep during the day, two participants in the no-expectation-wake condition took a nap.[9] There were no statistically significant differences between no-expectation-sleep and expectation-sleep conditions in (a) self-reported sleep duration, *t*(33) = 0.22, *p* = .829; (b) sleep duration assessed by fitness trackers, *t*(32) = 0.19, *p* = .847; (c) time awake assessed by fitness trackers, *t*(32) = 0.74, *p* = .465; (d) deep sleep assessed by fitness trackers, *t*(32) = 0.55, *p* = .588; (e) light sleep assessed by fitness trackers, *t*(32) = 0.40, *p* = .691; and (f) self-reported

---

[9] One participant napped for 30 min, the other for 1 hr. Note that comparable results emerged for memory performance and model-based analyses when excluding the two participants from the no-expectation-wake condition (see Tables S17 to S20 in the supplemental materials).

sleep quality, $t(32) = 0.37$, $p = .715$.[10] Specifically, the self-reported mean sleep duration for the night between learning and testing sessions was 7.42 hr (*SD* = 0.81) for participants in the no-expectation-sleep condition and 7.48 hr (*SD* = 0.65) for participants in the expectation-sleep condition. This corresponds to the mean sleep duration assessed by fitness trackers which was 7.82 hr (*SD* = 0.94; time awake: *M* = 0.01 hr, *SD* = 0.02; deep sleep: *M* = 1.94 hr, *SD* = 1.15; light sleep: *M* = 5.88 hr, *SD* = 1.15) in the no-expectation-sleep condition and 7.88 hr (*SD* = 0.74; time awake: *M* = 0.04 hr, *SD* = 0.14; deep sleep: *M* = 2.14 hr, *SD* = 0.95; light sleep: *M* = 5.74 hr, *SD* = 0.92) in the expectation-sleep condition. Participants in both sleep conditions reported medium sleep quality on a scale ranging from 1 (*very bad*) to 5 (*very good*; no expectation: *M* = 3.42, *SD* = 1.00; expectation: *M* = 3.55, *SD* = 0.96). To reach the learning criterion of 60%, participants repeated the sequence of stimulus presentation, distractor task, and immediate cued recall task on average 0.69 times (*SD* = 0.88). We observed no statistically significant differences in the number of repetitions between the four conditions, $F(3, 66) = 1.15$, $MSE = 0.76$, $p = .334$ (no expectation, wake: *M* = 0.59, *SD* = 0.80; no expectation, sleep: *M* = 0.46, *SD* = 0.66; expectation, wake: *M* = 1.00, *SD* = 0.77; expectation, sleep: *M* = 0.64, *SD* = 1.09). Furthermore, in their final immediate cued recall task, participants remembered on average 78% of the 40 words (*M* = 31.27 words, *SD* = 4.02), which is relatively close to the learning performance of 86% in the strong-encoding condition of Experiment 1 (strong encoding: *M* = 34.52 words, *SD* = 3.99; medium encoding: *M* = 22.67 words, *SD* = 6.80; see also Table 1).

Memory Performance Analysis

Because we manipulated test expectation at the end of the learning session following the immediate cued recall task, we did not expect immediate cued recall to differ between the four experimental conditions. As expected, a mixed ANOVA on immediate cued recall performance with test expectation and wake versus sleep as between-subjects factors revealed no statistically

---

[10] Note that in the expectation-sleep condition, sleep quality was not captured for one participant and the fitness tracker did not record any data for another participant.

significant differences, main effect of test expectation: $F(1, 66) = 2.34$, $MSE = 16.23$, $p = .131$, $\eta_p^2 = .03$; main effect of wake versus sleep: $F(1, 66) = 0.13$, $MSE = 16.23$, $p = .721$, $\eta_p^2 = .002$; interaction of test expectation and wake versus sleep: $F(1, 66) = 0.30$, $MSE = 16.23$, $p = .585$, $\eta_p^2 = .01$ (for the number of correctly recalled word pairs, see Table 4).

To test whether the sleep benefit is larger for expectation than no-expectation conditions as observed by Wilhelm et al. (2011), we used two-way ANOVAs on delayed cued recall, cued recall forgetting (i.e., the difference initial - delayed cued recall performance), and free recall as dependent variables with test expectation and wake versus sleep as between-subjects factors. The results showed no statistically significant main effect of test expectation for delayed cued recall, $F(1, 66) = 1.40$, $MSE = 30.70$, $p = .241$, $\eta_p^2 = .02$, cued recall forgetting, $F(1, 66) = 0.02$, $MSE = 6.55$, $p = .877$, $\eta_p^2 < .001$, and free recall, $F(1, 66) = 0.16$, $MSE = 23.58$, $p = .695$, $\eta_p^2 = .002$. In addition, we found a statistically significant main effect of wake versus sleep for delayed cued recall, $F(1, 66) = 5.41$, $MSE = 30.70$, $p = .023$, $\eta_p^2 = .08$, and cued recall forgetting, $F(1, 66) = 31.32$, $MSE = 6.55$, $p < .001$, $\eta_p^2 = .32$. Specifically, we observed better delayed cued recall performance and less forgetting after sleep than after wakefulness in both expectation and no-expectation conditions (see Table 4). Descriptively, free recall results also indicated a sleep benefit which was, however, not statistically significant, $F(1, 66) = 1.99$, $MSE = 23.58$, $p = .163$, $\eta_p^2 = .03$. The interaction of test expectation and retention interval was not statistically significant either; neither in delayed cued recall, $F(1, 66) = 0.37$, $MSE = 30.70$, $p = .547$, $\eta_p^2 = .01$, nor in cued recall forgetting, $F(1, 66) = 0.20$, $MSE = 6.55$, $p = .656$, $\eta_p^2 = .003$, nor in free recall, $F(1, 66) = 0.06$, $MSE = 23.58$, $p = .805$, $\eta_p^2 < .001$. Taken together, our results suggest sleep benefits for delayed cued recall, cued recall forgetting, and—to a lesser extent—free recall that are not affected by test expectation. These results are at odds with the results observed by Wilhelm et al. (2011).

Because we excluded four participants based on the application of the exclusion criteria recommended for the EMR model, we assessed the possible impact of these exclusions by performing all ANOVAs for the memory performance analysis again, this time including the previously excluded participants (i.e., new sample size: $n_{\text{no expectation/wake}} = 18$, $n_{\text{no expectation/sleep}} = 14$, $n_{\text{expectation/wake}} = 19$, $n_{\text{expectation/sleep}} = 23$). Very similar results emerged when using the larger sample size of $N = 74$, supporting the robustness of our findings (for details, see the supplemental materials).

Model-Based Analysis

As in Experiment 1, we first fitted the independence model, $G^2(20) = 47.29$, $p = .001$, followed by the 7-parameter EMR model with the relaxed independence assumption for parameter *u* ($u = u_s = u_u$), $G^2(16) = 34.85$, $p = .004$. Because both model versions produced misfit, we relaxed the independence assumption for parameter *s* ($s = s_s = s_u$) instead of parameter *u*, which also yielded misfit, $G^2(16) = 34.00$, $p = .005$. Finally, we relaxed the independence assumption for both parameter *s* ($s = s_s = s_u$) and parameter *u* ($u = u_s = u_u$). This 8-parameter model produced a slight misfit for the aggregated data, $G^2(12) = 21.50$, $p = .044$, that we consider still acceptable. The ML parameter estimates and standard errors are displayed in Table 5.

First, we checked whether the probability of successfully encoding word-pair associations (*e*) differed between the four experimental conditions. As expected, we observed no statistically significant differences in the encoding parameter *e*, $\Delta G^2(3) = 5.30$, $p = .151$. More importantly, participants encoded word pairs with a probability of about .79 irrespective of condition, which is relatively close to the encoding probability of .89 in the strong encoding condition of Experiment 1. Overall, the results show that, as expected, word-pair encoding success is unaffected by the test expectation manipulation that occurred after the immediate cued recall task.

Second, we tested the prediction of a larger sleep benefit in episodic memory storage for expectation than no-expectation conditions. For maintenance parameter *m*, we observed no statistically significant main effect of test expectation, $\Delta G^2(1) = 2.35$, $p = .125$. However, the main effect of wake versus sleep was statistically significant, $\Delta G^2(1) = 58.36$, $p < .001$, showing an approximately 10% larger probability estimate for the sleep than wake condition to successfully maintain encoded associations across the retention interval. By contrast, the interaction of test expectation and wake versus sleep was not statistically significant, $\Delta G^2(1) = 0.90$, $p = .343$. Taken together, these findings are clearly at odds with Wilhelm et al.'s (2011) prediction of a larger sleep benefit in episodic memory storage for test expectation.

Third, we tested whether the results in the strong-encoding condition for free-recall-retrieval parameter r_f of Experiment 1 can be replicated when applying a 60%-learning criterion. Consistent with the results obtained in Experiment 1, we found no statistically significant differences for the probability of successfully retrieving stored associations in free recall across the four experimental conditions, $\Delta G^2(3) = 2.89$, $p = .409$. Overall, the results thus support the hypothesis that the sleep benefit in episodic memory retrieval is reduced or might even vanish for higher levels of memory strength.

Fourth, we investigated whether cued recall retrieval (r_c) differs between the experimental conditions and found no significant differences, $\Delta G^2(3) = 2.78$, $p = .428$. In addition, participants retrieved stored associations with a probability of about .99 which corresponds to Experiment 1 and previous findings (see Erdfelder et al., 2022).

Finally, we checked the robustness of our results by performing the model-based analysis with Klauer's (2010) latent-trait approach implemented in TreeBUGS (Heck et al., 2018). Again, the estimated group-level means presented in Table B2 resemble those reported in Table 5, indicating that the result pattern is robust against complete versus partial pooling approaches to data analysis.

Discussion

One purpose of the current experiment was to investigate whether test expectation affects the sleep benefit in episodic memory storage. Wilhelm et al. (2011) showed that sleep compared to wakefulness improves memory retention performance (i.e., percentage of correctly recalled word pairs in the delayed cued recall test with the final-trial performance of the initial cued recall test set to 100%) in expectation, but not in no-expectation conditions. The authors proposed that relevant information is tagged for preferred memory consolidation during sleep, resulting in larger sleep benefits for memories that are expected to be tested in the future. We therefore predicted larger sleep-induced benefits for episodic memory storage in expectation than no-expectation conditions. Contrary to this prediction, our memory performance analyses suggest comparable sleep benefits in both expectation and no-expectation conditions for delayed cued recall, cued recall forgetting, and free recall. Thus, the sleep benefit in episodic memory appears not to be affected by test expectation. Although we based our minimum target sample size on the effect size of interest reported in Wilhelm et al. (2011; i.e., $d = 0.86$; $f = .43$), it may be that test-expectation effects on the sleep benefit are actually smaller than expected. Therefore, it could be that we were not able to detect the effect with a sample of 70 participants. To check this possibility, we conducted a post hoc sensitivity analysis with $N = 70$ participants and a conventional α-level of .05 for the ANOVA-based memory performance analysis. This sensitivity analysis confirmed that ANOVA $F(1, 66)$ tests can detect medium to large effects (i.e., $f = .34$) with a power of $1 - \beta = .80$ (Faul et al., 2007). Thus, even if the true effect is smaller than the results of Wilhelm et al. (2011) suggest, we should have been able to find test-expectation effects on the sleep benefit. Crucially, our results are in line with a recent study by Ashton and Cairney (2021) who also failed to replicate Wilhelm et al.'s (2011) finding even though larger sample sizes than in the current experiment were used (i.e., $N_{\text{Experiment 1}} = 162$, $N_{\text{Experiment 2}} = 124$). Specifically, Ashton and Cairney's (2021) study was conducted in an online setting using a 60% learning criterion for 40 (Experiment 1) and a 40% learning criterion for 100 semantically-related word pairs (Experiment 2). In both experiments, the authors observed better memory retention performance after sleep than wakefulness. However, neither a significant effect of test expectation, nor a significant interaction of test expectation and wake versus sleep emerged (Ashton & Cairney, 2021).[11] These results make it even less likely that our findings are simply the outcome of an underpowered experiment. In addition, Reverberi et al. (2020) casted further doubt

---

[11] Note that in Wilhelm et al.'s (2011) first experiment and in the two experiments by Ashton and Cairney (2021) the correct target word or word pair was presented after each response in the immediate cued recall task. Consequently, the observed differences between the experimental conditions in the delayed memory test could therefore also reflect additional learning following the final immediate cued recall trial. Because we offered no additional learning opportunity after completing the final trial, our memory performance measurement is not contaminated with additional learning influences.

on preferred memory consolidation of memories that are expected to be tested in the future. Specifically, the authors used two randomly mixed sets of picture-location associations (i.e., 80 building-location and 80 furniture-location associations) which were learned in three consecutive cycles of encoding and cued recall, resulting in an average of 71% correctly remembered associations. Prior to the 14.5-hr retention interval which included natural nighttime sleep, participants were informed that they would be tested and rewarded only for one of the two learned sets. Neither pre-post differences in cued recall nor functional magnetic resonance imaging (fMRI) data collected during cued recall after the retention interval showed significant effects of test expectation (Reverberi et al., 2020). Hence, our results are in line with recent findings suggesting that test expectation does not affect the sleep benefit in episodic memory, thereby challenging Wilhelm et al.’s (2011) results.

We also examined whether the results in the strong-encoding condition of Experiment 1 replicate when applying a 60%-learning criterion. In fact, for the maintenance parameter *m* our findings of a statistically significant sleep benefit matched those of Experiment 1. Most importantly, the results showed no significant sleep benefit in the free-recall-retrieval parameter for expectation and no-expectation conditions. Thus, we succeeded in replicating the reduced or rather even absent sleep benefit in episodic memory retrieval for strong memories found in Experiment 1.

General Discussion

Although it is a well-established finding that sleep compared to wakefulness fosters episodic memory, many open questions remain. One of these questions is whether and why salience or relevance of information affects how much episodic memories profit from sleep. Here, we tackled this question by investigating two salience cues—encoding strength and test expectation—that have been argued to tag information for preferred memory consolidation during or shortly after encoding. Using MPT modeling to take a more fine-graded look at the cognitive processes underlying the sleep benefit by disentangling storage and retrieval contributions to memory performance, we conducted two experiments:

In Experiment 1 we investigated tagging effects within the encoding phase by presenting half of the stimuli once versus three times. It is important to note that we manipulated encoding strength in the upper limb of Stickgold's (2009) inverted U-shaped function. For this range of encoding strengths, the consolidation account predicts that memory consolidation during sleep fosters medium-strength memories specifically. This prediction was supported by EEG (e.g., C. Schmidt et al., 2006) and TMR studies (e.g., Cairney et al., 2016; Creery et al., 2015). This notwithstanding, our behavioral results of Experiment 1 for memory storage were mixed. It rather appears that memory consolidation during sleep improves storage for all memories equally, irrespective of their encoding strength.

In Experiment 2, we examined tagging after encoding by manipulating whether a later memory test is announced after learning is completed. According to Wilhelm et al. (2011), a larger sleep benefit in storage should emerge when a later test was expected than when it was not. Contrary to this hypothesis but in line with other recent studies (Ashton & Cairney, 2021; Reverberi et al., 2020), we found no empirical evidence of a larger sleep benefit in episodic memory storage for stimuli expected to be tested in the future. Hence, together with the results of recent studies, our findings suggest that tagging relevant memories after encoding for preferred memory consolidation during sleep does not necessarily result in a larger sleep benefit—at least not for the salience cue "test expectation". In sum, behavioral results for salience-cue effects on episodic memory storage are mixed. One possible reason for the mixed findings is high variability in study design features (e.g., sleep study designs, salience cues) in this line of research. This makes it difficult to compare results across studies. Another reason is that effects of salience cues on the sleep benefit in episodic memory storage appear to be far smaller than previous studies suggest, which makes it hard to detect them. Most importantly, however, a better understanding of the tagging-mechanism and how it affects memory consolidation during sleep is needed to derive more concise hypotheses to evaluate the consolidation account.

Recent findings suggest that storage and retrieval might jointly contribute to the sleep benefit (Berres & Erdfelder, 2021; Erdfelder et al., 2022). In addition, previous research showed that retroactive interference decreases with higher study efforts (Howe, 1970; McGeoch, 1929). Taken together, these findings suggest an additional effect of encoding strength on sleep benefits in episodic memory retrieval. In line with this assumption, we observed a sleep benefit in episodic memory retrieval for medium- but not for strong-encoding conditions. The result in episodic memory retrieval for the strong-encoding condition of Experiment 1 was conceptually replicated in Experiment 2 by implementing a 60%-learning criterion.

However, why did a sleep benefit in episodic memory retrieval occur for medium encoding that was absent for strong encoding? As sketched above, previous studies showed that retroactive interference decreases with higher study efforts (Howe, 1970; McGeoch, 1929). Accordingly, with increasing memory strength, more items are successfully retrieved in both the sleep and the wake condition, leaving little room for further improvement of memory retrieval due to retroactive interference reduction during sleep. By implication, the sleep benefit in episodic memory retrieval should be reduced or might even vanish for high levels of memory strength. Overall, it thus appears that encoding strength affects sleep benefits in episodic memory retrieval more than sleep benefits in episodic memory storage.

Our objective in conducting this study was to advance our understanding of the cognitive processes underlying information salience effects on the sleep benefit in episodic memory. In fact, our findings have important implications for the interpretation and thus our understanding of sleep benefits. As sketched before, sleep benefits have typically been interpreted in terms of consolidation effects, neglecting interference reduction as an alternative source of sleep benefits (for a comprehensive review, see Berres & Erdfelder, 2021). Here we examined, for the first time, whether the salience cue “encoding strength” also affects the sleep benefit in episodic memory retrieval. By using the EMR model, we added further empirical evidence to the notion that two processes contribute to the sleep benefit, namely, memory consolidation and sleep-induced reduction of retroactive interference (cf. Berres & Erdfelder, 2021; Erdfelder et al., 2022). More importantly in the present context, we additionally showed that sleep benefits in episodic memory retrieval – but not those involved in episodic memory storage – are moderated by encoding strength. Apparently, memories of medium compared to high strength benefit more from the sleep-induced retrieval benefit.

Possible Limitations

As we compared natural nighttime sleep versus daytime wakefulness using a 12-hr retention interval in both experiments, the time of learning and testing differed between wake- and sleep conditions. Circadian effects could thus act as a possible confounding variable. Although we cannot rule out confounding effects completely, circadian effects are rather unlikely to occur in standard sleep study designs as used in our experiments. Past research that used additional morning and evening control conditions with short retention intervals (e.g., 20 min) has repeatedly shown comparable learning and test performance in either condition (e.g., Abel & Bäuml, 2012, 2013a, 2013b, 2014; Bäuml et al., 2014; Erdfelder et al., 2022; Fenn & Hambrick, 2013).

Moreover, the EMR model is based on an extended version of the free-then-cued-recall paradigm by Rouder and Batchelder (1998) where an immediate cued recall test right after learning is followed by a later free and cued recall test after the retention interval. Hence, immediate cued recall performance might in principle affect later free recall and delayed cued recall. According to the bifurcation model, retrieval per se should be seen as a learning event (e.g., Halamish & Bjork, 2011; Kornell et al., 2011; Rowland, 2014). Specifically, word pairs that were successfully retrieved during retrieval practice should have a greater memory strength than word pairs that were not successfully retrieved (Halamish & Bjork, 2011; Kornell et al., 2011). The bifurcation model thus predicts that, during retrieval practice, some word pairs are successfully retrieved and gain memory strength while other word pairs are not retrieved and remain at the same level of memory strength as before, leading to a bifurcation of the memory-strength distribution after the learning phase (Halamish & Bjork, 2011; Kornell et al., 2011; Rickard & Pan, 2018). For this reason, in a later memory test, more word pairs should exceed the retrieval threshold and thus be successfully retrieved when the immediate memory test was successful. This leads to the question whether the increased memory strength due to successful memory retrieval of some but not all word pairs in the immediate memory test biases our delayed free-recall- and delayed-cued-recall results. This appears unlikely for two reasons. First, immediate cued recall was used in all experimental conditions. Therefore, the memory-strength distribution should be bifurcated regardless of condition so that no condition benefits systematically. Second, the bifurcated memory-strength distribution might make it more difficult to detect sleep benefits in the upper limb of the inverted U-shaped function, especially when memory retrieval is relatively easy and thus most word pairs are above the retrieval threshold (e.g., delayed cued recall test; cf. Petzka et al., 2021). However, this does not affect the interpretation of our results per se. It thus is unlikely that our findings are systematically biased by effects of the initial cued recall test.

Outlook

Previous studies investigating the effects of salience cues on the sleep benefit in episodic memory differ not only in the salience manipulations employed but also in other features such as (a) sleep study designs (e.g., 1 hr nap vs. no nap, natural nighttime sleep vs. daytime wakefulness), (b) length of retention intervals (e.g., 4 hr, 12 hr), and (c) retrieval tasks (e.g., cued recall, free recall). This variability in study design features makes it difficult to compare results for a specific salience cue across studies. Therefore, future research should systematically investigate whether and how study design features moderate the sleep benefit in general and effects of information salience at encoding in particular.

In addition, more research is needed to examine how the sleep benefit and its underlying processes are affected by differences between encoding-strength manipulations. One way to manipulate encoding strength is to vary whether and how often stimuli are restudied or tested, and whether feedback is provided (e.g., Abel et al., 2019; Bäuml et al., 2014). Another way of manipulating encoding strength is to use different encoding instructions or to assess successfully applied deep encoding strategies (e.g., Bennion et al., 2016; Denis et al., 2020). Besides learning procedures, memory strength can also be manipulated by varying the emotionality of the stimulus material (e.g., Hu et al., 2006; Payne et al., 2015). As such, emotional stimuli receive more attention and also more elaboration than neutral stimuli at encoding (Hamann, 2001). One caveat of these encoding-quality manipulations is that it is difficult to compare memory performance results across studies in terms of memory strength. In addition, these procedures do not allow for a reliable fine-grained memory-strength manipulation which is essential for a more rigorous test of Stickgold's (2009) inverted U-shaped function. We therefore suggest using better controllable perceptual stimuli (e.g., by manipulating brightness contrast) and to systematically vary encoding strength in small steps across the full possible range by employing the same strength manipulation.

Conclusion

Using MPT modeling to disentangle encoding, storage, and retrieval, we examined whether and how cognitive processes underlying the sleep benefit in episodic memory are affected by two salience cues—encoding strength and test expectation. We showed that encoding strength affects retrieval more than storage contributions to the sleep benefit. In addition, both storage and retrieval contributions to the sleep benefit were unaffected by test expectation. Although further research is needed, we provide first evidence that retroactive interference under wakefulness is necessary for the sleep benefit in memory retrieval to occur. If retroactive interference effects are suppressed by very high memory strength even under wakefulness (cf. Howe, 1970; McGeoch, 1929), there is simply no basis for a sleep-induced reduction of retroactive interference. By implication, the sleep benefit in memory retrieval reduces or even vanishes when memory strength approaches the ceiling. Hence, our findings have important theoretical implications because they indicate that variations of the sleep benefit in episodic memory may not always be due to variations in storage but at least sometimes also be due to variations in retrieval success.

References

Abel, M., & Bäuml, K.-H. T. (2012). Retrieval-induced forgetting, delay, and sleep. *Memory, 20*(5), 420–428. https://doi.org/10.1080/09658211.2012.671832

Abel, M., & Bäuml, K.-H. T. (2013a). Adaptive memory: The influence of sleep and wake delay on the survival-processing effect. *Journal of Cognitive Psychology, 25*(8), 917–924. https://doi.org/10.1080/20445911.2013.825621

Abel, M., & Bäuml, K.-H. T. (2013b). Sleep can eliminate list-method directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(3), 946–952. https://doi.org/10.1037/a0030529

Abel, M., & Bäuml, K.-H. T. (2014). Sleep can reduce proactive interference. *Memory, 22*(4), 332–339. https://doi.org/10.1080/09658211.2013.785570

Abel, M., Haller, V., Köck, H., Pötschke, S., Heib, D., Schabus, M., & Bäuml, K.-H. T. (2019). Sleep reduces the testing effect—But not after corrective feedback and prolonged retention interval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(2), 272–287. https://doi.org/10.1037/xlm0000576

Antony, J. W., & Paller, K. A. (2017). Hippocampal contributions to declarative memory consolidation during sleep. In D. F. Hannula & M. C. Duff (Eds.), *The hippocampus from cells to systems* (1st ed., pp. 245–280). Springer International Publishing. https://doi.org/10.1007/978-3-319-50406-3

Ashton, J. E., & Cairney, S. A. (2021). Future-relevant memories are not selectively strengthened during sleep. *PLOS ONE, 16*(11), Article e0258110. https://doi.org/10.1371/journal.pone.0258110

Backhaus, J., & Junghanns, K. (2006). Daytime naps improve procedural motor memory. *Sleep Medicine*, *7*(6), 508–512. https://doi.org/10.1016/j.sleep.2006.04.002

Bäuml, K.-H. T., Holterman, C., & Abel, M. (2014). Sleep can reduce the testing effect: It enhances recall of restudied items but can leave recall of retrieved items unaffected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1568–1581. https://doi.org/10.1037/xlm0000025

Baran, B., Daniels, D., & Spencer, R. M. C. (2013). Sleep-dependent consolidation of value-based learning. *PLOS ONE*, *8*(10), Article e75326. https://doi.org/10.1371/journal.pone.0075326

Barco, A., Lopez de Armentia, M., & Alarcon, J. M. (2008). Synapse-specific stabilization of plasticity processes: The synaptic tagging and capture hypothesis revisited 10 years later. *Neuroscience & Biobehavioral Reviews*, *32*(4), 831–851. https://doi.org/10.1016/j.neubiorev.2008.01.002

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*(1), 57–86. https://doi.org/10.3758/BF03210812

Bennion, K. A., Payne, J. D., & Kensinger, E. A. (2016). The impact of napping on memory for future-relevant stimuli: Prioritization among multiple salience cues. *Behavioral Neuroscience, 130*(3), 281–289. https://doi.org/10.1037/bne0000142

Berres, S., & Erdfelder, E. (2021). The sleep benefit in episodic memory: An integrative review and a meta-analysis. *Psychological Bulletin, 147*(12), 1309–1353. https://doi.org/10.1037/bul0000350

Buzsáki, G. (1998). Memory consolidation during sleep: A neurophysiological perspective. *Journal of Sleep Research*, *7*(S1), 17–23. https://doi.org/10.1046/j.1365-2869.7.s1.3.x

Cairney, S. A., Lindsay, S., Sobczak, J. M., Paller, K. A., & Gaskell, M. G. (2016). The benefits of targeted memory reactivation for consolidation in sleep are contingent on memory accuracy and direct cue-memory associations. *Sleep, 39*(5), 1139–1150. https://doi.org/10.5665/sleep.5772

Creery, J. D., Oudiette, D., Antony, J. W., & Paller, K. A. (2015). Targeted memory reactivation during sleep depends on prior learning. *Sleep, 38*(5), 755–763. https://doi.org/10.5665/sleep.4670

Denis, D., Mylonas, D., Poskanzer, C., Bursal, V., Payne, J. D., & Stickgold, R. (2021). Sleep spindles preferentially consolidate weakly encoded memories. *Journal of Neuroscience, 41*(18), 4088–4099. https://doi.org/10.1523/JNEUROSCI.0818-20.2021

Denis, D., Schapiro, A. C., Poskanzer, C., Bursal, V., Charon, L., Morgan, A., & Stickgold, R. (2020). The role of item exposure and visualization success in the consolidation of memories across wake and sleep. *Learning & Memory, 27*, 451–456. https://doi.org/10.1101/lm.051383.120

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

Drosopoulos, S., Schulze, C., Fischer, S., & Born, J. (2007). Sleep's function in the spontaneous recovery and consolidation of memories. *Journal of Experimental Psychology: General*, *136*(2), 169–183. https://doi.org/10.1037/0096-3445.136.2.169

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annual Review of Psychology*, *55*, 51–86. https://doi.org/10.1146/annurev.psych.55.090902.142050

Dudai, Y. (2012). The restless engram: Consolidations never end. *Annual Review of Neuroscience*, *35*, 227–247. https://doi.org/10.1146/annurev-neuro-062111-150500

Dudai, Y., Karni, A., & Born, J. (2015). The consolidation and transformation of memory. *Neuron*, *88*(1), 20–32. https://doi.org/10.1016/j.neuron.2015.09.004

Dyne, A. M., Humphreys, M. S., Bain, J. D., & Pike, R. (1990). Associative interference effects in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(5), 813–824. https://doi.org/10.1037/0278-7393.16.5.813

Ecker, U. K. H., Brown, G. D. A., & Lewandowsky, S. (2015). Memory without consolidation: Temporal distinctiveness explains retroactive interference. *Cognitive Science*, *39*(7), 1570–1593. https://doi.org/10.1111/cogs.12214

Ecker, U. K. H., Tay, J.-X., & Brown, G. D. A. (2015). Effects of prestudy and poststudy rest on memory: Support for temporal interference accounts of forgetting. *Psychonomic Bulletin & Review*, *22*, 772–778. https://doi.org/10.3758/s13423-014-0737-8

Eckes, T. (1994). Zur internen Struktur semantischer Kategorien: Typikalitätsnormen auf der Basis von Ratings [On the internal structure of semantic categories: Typicality norms on the basis of ratings]. In W. Hager & M. Hasselhorn (Eds.), *Handbuch deutschsprachiger Wortnormen* (pp. 128–137). Hogrefe. (Reprinted from „Zur internen Struktur semantischer Kategorien: Typikalitätsnormen auf der Basis von Ratings“, 1985, *Sprache & Kognition, 4*[4], 192–202)

Ekstrand, B. R. (1967). Effect of sleep on memory. *Journal of Experimental Psychology*, *75*(1), 64–72. https://doi.org/10.1037/h0024907

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie*, *217*(3), 108–124. https://doi.org/10.1027/0044-3409.217.3.108

Erdfelder, E., Berres, S., Quevedo Pütter, J., & Küpper-Tetzel, C. E. (2022). *Why does sleep improve episodic memory? An encoding-maintenance-retrieval analysis* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Feld, G. B., Lange, T., Gais, S., & Born, J. (2013). Sleep-dependent declarative memory consolidation—Unaffected after blocking NMDA or AMPA receptors but enhanced by NMDA coagonist D-cycloserine. *Neuropsychopharmacology*, *38*, 2688–2697. https://doi.org/10.1038/npp.2013.179

Fenn, K. M., & Hambrick, D. Z. (2012). Individual differences in working memory capacity predict sleep-dependent memory consolidation. *Journal of Experimental Psychology: General*, *141*(3), 404–410. https://doi.org/10.1037/a0025268

Fenn, K. M., & Hambrick, D. Z. (2013). What drives sleep-dependent memory consolidation: Greater gain or less loss? *Psychonomic Bulletin & Review*, *20*, 501–506. https://doi.org/10.3758/s13423-012-0366-z

Fernandez, L. M. J., & Lüthi, A. (2020). Sleep spindles: Mechanisms and functions. *Physiological Reviews, 100*(2), 805–868. https://doi.org/10.1152/physrev.00042.2018

Flammer, A., Reisbeck, C., & Stadler, S. (1985). Typikalitätsnormen für dreizehn Begriffe in einer deutschschweizerischen Studentenpopulation [Norms of typicality for thirteen concepts in a population of Swiss-German students]. *Sprache & Kognition, 4*(1), 49–63.

Frey, U., & Morris, R. G. M. (1998). Synaptic tagging: Implications for late maintenance of hippocampal long-term potentiation. *Trends in Neurosciences*, *21*(5), 181–188. https://doi.org/10.1016/S0166-2236(97)01189-2

Gais, S., & Born, J. (2004). Low acetylcholine during slow-wave sleep is critical for declarative memory consolidation. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(7), 2140–2144. https://doi.org/10.1073/pnas.0305404101

Gelman, A., & Rubin, D. B. (1992). Interference from iterative simulation using multiple sequences. *Statistical Science, 7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 801–812. https://doi.org/10.1037/a0023219

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Sciences*, *5*(9), 394–400. https://doi.org/10.1016/S1364-6613(00)01707-1

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods, 50*, 264–284. https://doi.org/10.3758/s13428-017-0869-7

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

Howe, T. S. (1970). Joint effects of proactive and retroactive interference as a function of degree of learning. *Journal of Experimental Psychology, 83*(1), 68–72. https://doi.org/10.1037/h0028553

Hu, P., Stylos-Allan, M., & Walker, M. P. (2006). Sleep facilitates consolidation of emotional declarative memory. *Psychological Science*, *17*(10), 891–898. https://doi.org/10.1111/j.1467-9280.2006.01799.x

Jenkins, J. G., & Dallenbach, K. M. (1924). Obliviscence during sleep and waking. *The American Journal of Psychology*, *35*(4), 605–612. https://doi.org/10.2307/1414040

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika, 75*(1), 70–98. https://doi.org/10.1007/s11336-009-9141-0

Klinzing, J. G., Niethard, N., & Born, J. (2019). Mechanisms of systems memory consolidation during sleep. *Nature Neuroscience*, *22*(10), 1598–1610. https://doi.org/10.1038/s41593-019-0467-3

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, *65*(2), 85–97. https://doi.org/10.1016/j.jml.2011.04.002

Küpper-Tetzel, C. E., & Erdfelder, E. (2012). Encoding, maintenance, and retrieval processes in the lag effect: A multinomial processing tree analysis. *Memory, 20*(1), 37–47. https://doi.org/10.1080/09658211.2011.631550

Küpper-Tetzel, C. E., Erdfelder, E., & Dickhäuser, O. (2014). The lag effect in secondary school classrooms: Enhancing students' memory for vocabulary. *Instructional Science, 42*, 373–388. https://doi.org/10.1007/s11251-013-9285-2

Kuhlmann, B. G., Erdfelder, E., & Moshagen, M. (2019). Testing interactions in multinomial processing tree models. *Frontiers in Psychology, 10*, Article 2364. https://doi.org/10.3389/fpsyg.2019.02364

Leiner, D. J. (2018). *SoSci Survey* (Version 2.5.04) [Computer software]. https://www.soscisurvey.de

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE Publications.

Mannhaupt, H.-R. (1983). Produktionsnormen für verbale Reaktionen zu 40 geläufigen Kategorien [German category norms for verbal items in 40 categories]. *Sprache & Kognition, 2*(4), 264–278.

Marshall, L., Mölle, M., Hallschmid, M., & Born, J. (2004). Transcranial direct current stimulation during sleep improves declarative memory. *Journal of Neuroscience*, *24*(44), 9985–9992. https://doi.org/10.1523/JNEUROSCI.2725-04.2004

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

McGeoch, J. A. (1929). The influence of degree of learning upon retroactive inhibition. *The American Journal of Psychology, 41*(2), 252–262. https://doi.org/10.2307/1415236

McKinney, F. (1935). Retroactive inhibition and recognition memory. *Journal of Experimental Psychology, 18*(5), 585–598. https://doi.org/10.1037/h0054474

Morris, R. G. M. (2006). Elements of a neurobiological theory of hippocampal function: The role of synaptic plasticity, synaptic tagging and schemas. *European Journal of Neuroscience*, *23*(11), 2829–2846. https://doi.org/10.1111/j.1460-9568.2006.04888.x

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods, 42*(1), 42–54. https://doi.org/10.3758/BRM.42.1.42

Muehlroth, B. E., Rasch, B., & Werkle-Bergner, M. (2020). Episodic memory consolidation during sleep in healthy aging. *Sleep Medicine Reviews, 52*, Article 101304. https://doi.org/10.1016/j.smrv.2020.101304

Oudiette, D., Antony, J. W., Creery, J. D., & Paller, K. A. (2013). The role of memory reactivation during wakefulness and sleep in determining which memories endure. *Journal of Neuroscience*, *33*(15), 6672–6678. https://doi.org/10.1523/JNEUROSCI.5497-12.2013

Payne, J. D., Kensinger, E. A., Wamsley, E. J., Spreng, R. N., Alger, S. E., Gibler, K., Schacter, D. L., & Stickgold, R. (2015). Napping and the selective consolidation of negative aspects of scenes. *Emotion*, *15*(2), 176–186. https://doi.org/10.1037/a0038683

Petzka, M., Charest, I., Balanos, G. M., & Staresina, B. P. (2021). Does sleep-dependent consolidation favour weak memories? *Cortex, 134*, 65–75. https://doi.org/10.1016/j.cortex.2020.10.005

Peyrache, A., & Seibt, J. (2020). A mechanism for learning with sleep spindles. *Philosophical Transactions of the Royal Society B, 375*(1799). Article 20190230. https://doi.org/10.1098/rstb.2019.0230

Plihal, W., & Born, J. (1997). Effects of early and late nocturnal sleep on declarative and procedural memory. *Journal of Cognitive Neuroscience*, *9*(4), 534–547. https://doi.org/10.1162/jocn.1997.9.4.534

Plummer, M. (2003, March 20–22). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling* [Paper presentation]. Third International Workshop on Distributed Statistical Computing, Vienna, Austria. http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Drafts/Plummer.pdf

Postman, L. (1952). Retroactive inhibition in recall and recognition. *Journal of Experimental Psychology, 44*(3), 165–169. https://doi.org/10.1037/h0057388

Quevedo Pütter, J., & Erdfelder, E. (2022). Alcohol-induced retrograde facilitation? Mixed evidence in a preregistered replication and encoding-maintenance-retrieval analysis. *Experimental Psychology, 69*(6), 335–350. https://doi.org/10.1027/1618-3169/a000569

Randler, C. (2013). German version of the reduced Morningness-Eveningness Questionnaire (rMEQ). *Biological Rhythm Research, 44*(5), 730–736. https://doi.org/10.1080/09291016.2012.739930

Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Reviews*, *93*(2), 681–766. https://doi.org/10.1152/physrev.00032.2012

Rauchs, G., Feyers, D., Landeau, B., Bastin, C., Luxen, A., Maquet, P., & Collette, F. (2011). Sleep contributes to the strengthening of some memories over others, depending on hippocampal activity at learning. *Journal of Neuroscience*, *31*(7), 2563–2568. https://doi.org/10.1523/JNEUROSCI.3972-10.2011

Redondo, R. L., & Morris, R. G. M. (2011). Making memories last: The synaptic tagging and capture hypothesis. *Nature Reviews Neuroscience*, *12*, 17–30. https://doi.org/10.1038/nrn2963

Reverberi, S., Kohn, N., & Fernández, G. (2020). No evidence for an effect of explicit relevance instruction on consolidation of associative memories. *Neuropsychologia, 143*, Article 107491. https://doi.org/10.1016/j.neuropsychologia.2020.107491

Rickard, T. C., & Pan, S. C. (2018). A dual memory theory of the testing effect. *Psychonomic Bulletin & Review*, *25*, 847–869. https://doi.org/10.3758/s13423-017-1298-4

Rohatgi, A. (2019). *WebPlotDigitizer—Web based plot digitizer* (Version 4.1, 4.2) [Web-based computer software]. https://automeris.io/We bPlotDigitizer/

Rouder, J. N., & Batchelder, W. H. (1998). Multinomial models for measuring storage and retrieval processes in paired associate learning. In C. E. Dowling, F. S. Roberts, & P. Theuns (Eds.), *Recent progress in mathematical psychology: Psychophysics, knowledge, representation, cognition, and measurement* (pp. 195–225). Lawrence Erlbaum Associates Publishers.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140*(6), 1432–1463. https://doi.org/10.1037/a0037559

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. https://doi.org/10.1093/cercor/bhr034

Scheithe, K., & Bäuml, K.-H. (1995). Deutschsprachige Normen für Vertreter von 48 Kategorien. *Sprache & Kognition, 14*(1), 39–43.

Schmidt, C., Peigneux, P., Muto, V., Schenkel, M., Knoblauch, V., Münch, M., de Quervain, D. J.-F., Wirz-Justice, A., & Cajochen, C. (2006). Encoding difficulty promotes postlearning changes in sleep spindle activity during napping. *The Journal of Neuroscience, 26*(35), 8976–8982. https://doi.org/10.1523/JNEUROSCI.2464-06.2006

Schmidt, O., Erdfelder, E., & Heck, D. W. (2023). How to develop, test, and extend multinomial processing tree models: A tutorial. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000561

Schoch, S. F., Cordi, M. J., & Rasch, B. (2017). Modulating influences of memory strength and sensitivity of the retrieval test on the detectability of the sleep consolidation effect. *Neurobiology of Learning and Memory*, *145*, 181–189. https://doi.org/10.1016/j.nlm.2017.10.009

Sheth, B. R., Varghese, R., & Truong, T. (2012). Sleep shelters verbal memory from different kinds of interference. *Sleep*, *35*(7), 985–996. https://doi.org/10.5665/sleep.1966

Stickgold, R. (2009). How do I remember? Let me count the ways. *Sleep Medicine Reviews*, *13*(5), 305–308. https://doi.org/10.1016/j.smrv.2009.05.004

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

Tucker, M. A., & Fishbein, W. (2008). Enhancement of declarative memory performance following a daytime nap is contingent on strength of initial task acquisition. *Sleep, 31*(2), 197–203. https://doi.org/10.1093/sleep/31.2.197

Tulving, E., & Psotka, J. (1971). Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. *Journal of Experimental Psychology*, *87*(1), 1–8. https://doi.org/10.1037/h0030185

Ukraintseva, Y. V, & Dorokhov, V. B. (2012). Effects of daytime sleep on the consolidation of declarative memory in humans. *Neuroscience and Behavioral Physiology*, *42*, 700–706. https://doi.org/10.1007/s11055-012-9621-z

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

Yonelinas, A. P., Ranganath, C., Ekstrom, A. D., & Wiltgen, B. J. (2019). A contextual binding theory of episodic memory: Systems consolidation reconsidered. *Nature Reviews Neuroscience*, *20*, 364–375. https://doi.org/10.1038/s41583-019-0150-4

Table 1

Study Characteristics and Percentage of Correct Responses Averaged Across Sleep and Wake Conditions in Immediate Memory Tests of Relevant Experiments

References	Sample size	Sleep study design	Immediate memory test	Percentage of correct immediate test responses		Sleep benefit *d* [95% CI] [b]
				Lower limb (≲ 50%)	Upper limb (≳ 50%)	
		Previous research				
Cairney et al., 2016; picture-word task	30	Daytime nap	Cued recall			
“Weak” condition					65.20	—
“Strong” condition					85.60	—
Denis et al., 2020; not visualized, 12-hr interval [a]	39	Natural sleep and wakefulness	Cued recall			
“Weak” condition				27.11		0.43 [-0.21, 1.06] [a]
“Intermediate” condition					53.81	0.29 [-0.34, 0.92] [a]
“Strong” condition					66.88	0.70 [-0.05, 1.35] [a]
Denis et al., 2020; visualized, 12-hr interval [a]	39	Natural sleep and wakefulness	Cued recall			
“Weak” condition				38.48		1.52 [0.81, 2.23] [a]
“Intermediate” condition					64.74	1.28 [0.59, 1.97] [a]
“Strong” condition					84.22	1.00 [0.34, 1.67] [a]
Denis et al., 2020; overall, 12-hr interval	39	Natural sleep and wakefulness	Cued recall			
“Weak” condition				32.48		0.72 [0.08, 1.37]
“Intermediate” condition					63.33	0.72 [0.07, 1.37]
“Strong” condition					81.19	0.51 [-0.13, 1.15]

References	Sample size	Sleep study design	Immediate memory test	Percentage of correct immediate test responses		Sleep benefit *d* [95% CI] [b]
				Lower limb (≲ 50%)	Upper limb (≳ 50%)	
Denis et al., 2021	54	Daytime nap	Cued recall			
“Weak” condition				31.46		0.23 [-0.34, 0.80]
“Intermediate” condition					58.25	0.06 [-0.50, 0.63]
“Strong” condition					77.26	-0.04 [-0.60, 0.53]
Drosopoulos et al., 2007; experiment 2, overall	40	Total sleep deprivation in the first night after learning with recovery night	Cued recall			
“Weak” condition					82.50	0.75 [-0.16, 1.65]
“Intense” condition					95.88	-0.13 [-1.01, 0.75]
Petzka et al., 2021; sequence memory, no interference	60	Natural sleep and wakefulness	Cued recall			
“Weak” condition					66.32	0.96 [0.21, 1.72]
“Strong” condition					91.58	-0.20 [-0.92, 0.51]
Petzka et al., 2021; sequence memory, interference	60	Natural sleep and wakefulness	Cued recall			
“Weak” condition					63.51	0.60 [-0.13, 1.34]
“Strong” condition					93.86	1.61 [0.79, 2.44]
C. Schmidt et al., 2006	13	Daytime nap	Cued recall			
“Difficult” condition				48.30		—
“Easy” condition					66.00	—
Schoch et al., 2017; group 2	57	Natural sleep and wakefulness	Free recall (image description using keywords)			
“Neutral images” condition				27.19		-0.32 [-0.85, 0.20]
“Emotional images” condition				45.10		0.31 [-0.21, 0.83]

References	Sample size	Sleep study design	Immediate memory test	Percentage of correct immediate test responses		Sleep benefit *d* [95% CI] [b]
				Lower limb (≲ 50%)	Upper limb (≳ 50%)	
Sheth et al., 2012; experiment 2, 12-hr interval	62	Natural sleep and wakefulness	Cued recall			
“1-time correct recall” condition					100.00	3.02 [1.74, 4.30]
“3-times correct recall” condition					100.00	1.30 [0.37, 2.22]
“6-times correct recall” condition					100.00	-0.31 [-1.19, 0.57]
Tucker & Fishbein, 2008; paired associates	33	Daytime nap	Cued recall			
“No immediate test” condition				—		-0.09 [-0.78, 0.59] [a]
“Immediate test” condition				42.41		0.84 [0.13, 1.55] [a]
Ukraintseva & Dorokhov, 2012	14	Daytime nap	Cued recall			
“1-time memorization” condition				—		0.70 [-0.07, 1.46] [a]
“2-times memorization” condition				47.87		0.22 [-0.52, 0.97] [a]
		Current experiments				
Experiment 1	27	Natural sleep and wakefulness	Cued recall			
“Medium” condition, cued recall					56.66	0.20 [-0.56, 0.96]
“Strong” condition, cued recall					86.30	0.72 [-0.06, 1.50]
“Medium” condition, free recall					56.66	0.79 [0.01, 1.58]
“Strong” condition, free recall					86.30	0.05 [-0.71, 0.81]
Experiment 2	70	Natural sleep and wakefulness	Cued recall			
“No test expectation” condition, cued recall					76.08	1.49 [0.68, 2.31]
“Test expectation” condition, cued recall					79.75	0.81 [0.16, 1.45]
“No test expectation” condition, free recall					76.08	0.78 [0.03, 1.53]
“Test expectation” condition, free recall					79.75	0.59 [-0.05, 1.22]

Note. The reported percentage of correct immediate test responses is averaged across sleep and wake conditions. We calculated Cohen's *d* on the basis of the reported means and standard deviations or standard errors. Note that positive and negative *d*-values indicate a sleep benefit and a sleep disadvantage compared to wakefulness in episodic memory, respectively. Because the studies by Cairney et al. (2016), Creery et al. (2015), and C. Schmidt et al. (2006) included no wake condition, no Cohen's *d* was computed. Moreover, three relevant studies reported also other memory performance measures: First, Cairney et al. (2016) conducted a picture-location task with 30 participants and reported location errors of 3.54 cm for the "weak" and 2.19 cm for the "strong" condition in the immediate cued recall task. Second, Creery et al. (2015) conducted a picture-location task with 20 participants and reported location errors of 99.25 px (2.63 cm) for bottom-half learners and 68.70 px (1.82 cm) for top-half learners in the immediate cued recall task. Finally, Petzka et al. (2021) conducted a spatial memory test with 60 participants and reported error rates in the immediate cued recall task of 56.78% (weak encoding) versus 33.49% (strong encoding) for no-interference conditions (weak encoding: *d* = 0.09, 95% CI [-0.63, 0.81]; strong encoding: *d* = -0.01, 95% CI [-0.73, 0.70]) and of 66.52% (weak encoding) versus 33.17% (strong encoding) for interference conditions (weak encoding: *d* = 0.54, 95% CI [-0.19, 1.27]; strong encoding: *d* = 1.28, 95% CI [0.50, 2.07]), averaged across sleep and wake conditions.

[a] We used the open-source WebPlotDigitizer (Rohatgi, 2019) to obtain the necessary values for calculating the mean percentage of correct immediate test responses or Cohen's *d*.

[b] The sleep benefit effect size estimates (*d*) and the respective 95% confidence intervals were calculated with the web-based effect size calculator provided by David B. Wilson (https://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php; Lipsey & Wilson, 2001)

Table 2

Means and Standard Errors for the Memory Performance Measures in Experiment 1

Dependent variable	Medium encoding		Strong encoding	
	Wake	Sleep	Wake	Sleep
Immediate cued recall	22.93 (1.38)	22.33 (1.27)	34.27 (0.87)	34.83 (0.65)
Delayed cued recall	20.73 (1.43)	21.75 (1.32)	31.93 (1.08)	34.58 (0.84)
Cued recall forgetting	2.20 (0.26)	0.58 (0.29)	2.33 (0.57)	0.25 (0.36)
Free recall	5.80 (0.83)	8.08 (0.69)	14.53 (1.32)	14.75 (0.80)

Note. The table shows the mean number of words (cued recall) and word pairs (free recall) correctly recalled in the respective memory test. Standard errors of the mean are presented in parentheses. Cued recall forgetting represents the difference between immediate and delayed cued recall performance. Participants were randomly assigned to a wake (n = 15) and sleep (n = 12) condition.

Table 3

Maximum-Likelihood Parameter Estimates and Standard Errors of the Encoding-Maintenance-Retrieval (EMR) Multinomial Model for Experiment 1

Parameter	Medium encoding		Strong encoding	
	Wake	Sleep	Wake	Sleep
e	**.59 (.02)**	**.57 (.02)**	**.88 (.01)**	**.89 (.01)**
m	**.90 (.02)**	**.97 (.01)**	**.93 (.01)**	**.99 (.01)**
r_c	.98 (.01)	.98 (.01)	.98 (.01)	.99 (.01)
$\boldsymbol{r_f}$	**.27 (.03)**	**.36 (.03)**	**.44 (.02)**	**.42 (.02)**
s	.06 (.01)	.04 (.01)	.09 (.01)	.03 (.01)
u_s	.05 (.03)	.22 (.11)	.19 (.05)	.23 (.19)
u_u	.03 (.01)	.06 (.01)	.03 (.01)	.09 (.03)

Note. The table shows the estimated success probabilities of latent cognitive processes. Standard errors of the parameter estimates are presented in parentheses. Estimation results for core parameters are displayed in bold. e = probability of successful encoding of word-pair associations; m = probability of maintaining encoded associations across the retention interval; r_c = probability of retrieving stored associations in cued recall; r_f = probability of retrieving stored associations in free recall; s = probability of associated single word retrieval during free recall; u_s = probability of non-associated single word retrieval during free recall for successful immediate cued-recall retrieval; u_u = probability of non-associated single word retrieval during free recall for unsuccessful immediate cued-recall retrieval.

Table 4

Means and Standard Errors for the Memory Performance Measures in Experiment 2

Dependent variable	No test expectation		Test expectation	
	Wake (*n* = 17)	Sleep (*n* = 13)	Wake (*n* = 18)	Sleep (*n* = 22)
Immediate cued recall	30.35 (0.48)	30.54 (0.50)	32.39 (0.56)	31.50 (0.39)
Delayed cued recall	25.59 (0.73)	29.54 (0.57)	28.00 (0.80)	30.32 (0.51)
Cued recall forgetting	4.76 (0.42)	1.00 (0.26)	4.39 (0.31)	1.18 (0.20)
Free recall	14.12 (0.62)	16.08 (0.67)	13.94 (0.64)	15.32 (0.43)

Note. The table shows the mean number of words (cued recall) and word pairs (free recall) correctly recalled in the respective memory test. Standard errors of the mean are presented in parentheses. Cued recall forgetting represents the difference between immediate and delayed cued recall performance.

Table 5

Maximum-Likelihood Parameter Estimates and Standard Errors of the Encoding-Maintenance-Retrieval (EMR) Multinomial Model for Experiment 2

Parameter	No test expectation		Test expectation	
	Wake	Sleep	Wake	Sleep
e	**.77 (.02)**	**.78 (.02)**	**.82 (.01)**	**.80 (.01)**
m	**.85 (.02)**	**.96 (.01)**	**.87 (.01)**	**.96 (.01)**
r_c	.99 (.01)	.98 (.01)	.99 (.01)	.99 (.01)
r_f	**.53 (.02)**	**.53 (.03)**	**.48 (.02)**	**.50 (.02)**
s_s	.10 (.02)	.06 (.01)	.10 (.01)	.06 (.01)
s_u	.41 (.22)	.51 (.19)	.20 (.19)	.00 (.49)
u_s	.15 (.03)	.16 (.08)	.23 (.04)	.15 (.05)
u_u	.10 (.02)	.13 (.02)	.11 (.02)	.13 (.02)

Note. The table shows the estimated success probabilities of latent cognitive processes. Standard errors of the parameter estimates are presented in parentheses. Estimation results for core parameters are displayed in bold. e = probability of successful encoding of word-pair associations; m = probability of maintaining encoded associations across the retention interval; r_c = probability of retrieving stored associations in cued recall; r_f = probability of retrieving stored associations in free recall; s_s = probability of associated single word retrieval during free recall for successful immediate cued-recall retrieval; s_u = probability of associated single word retrieval during free recall for unsuccessful immediate cued-recall retrieval; u_s = probability of non-associated single word retrieval during free recall for successful immediate cued-recall retrieval; u_u = probability of non-associated single word retrieval during free recall for unsuccessful immediate cued-recall retrieval.

Appendix A

The Encoding-Maintenance-Retrieval (EMR) Multinomial Model

The Encoding-Maintenance-Retrieval (EMR) model is a multinomial model for categorical data. Its parameters represent success probabilities of latent cognitive processes involved in an extended version of the free-then-cued-recall paradigm by Rouder and Batchelder (1998). In the extended paradigm, participants study a set of weakly related word pairs (e.g., “candy – bread”), followed by three memory tests: First, an immediate cued recall test right after learning in which participants are provided with the first word of the word pair as a cue and have to recall the second word (e.g., “candy – ??”). Second, a free recall test after a longer retention interval—filled with either sleep or wakefulness—in which participants are instructed to recall all word pairs and single words they can remember. Third, a delayed cued recall test after finishing the free recall test, resembling the immediate cued recall test. For each studied word pair, there are two possible outcomes in cued recall (i.e., correct vs. incorrect) and three possible outcomes in free recall tests (i.e., both words recalled, one word recalled, no word recalled). Consequently, there are $2 \cdot 3 \cdot 2 = 12$ possible outcome patterns per word pair. The resulting observation categories E_1 to E_{12} and their definitions are summarized in Table A1.

Table A1

Observation Categories for the Extended Free-Then-Cued-Recall Paradigm

Immediate cued recall	Free recall			Delayed cued recall
	Both words	One word	No word	
Correct	E_1	E_2	E_3	Correct
	E_4	E_5	E_6	Incorrect
Incorrect	E_7	E_8	E_9	Correct
	E_{10}	E_{11}	E_{12}	Incorrect

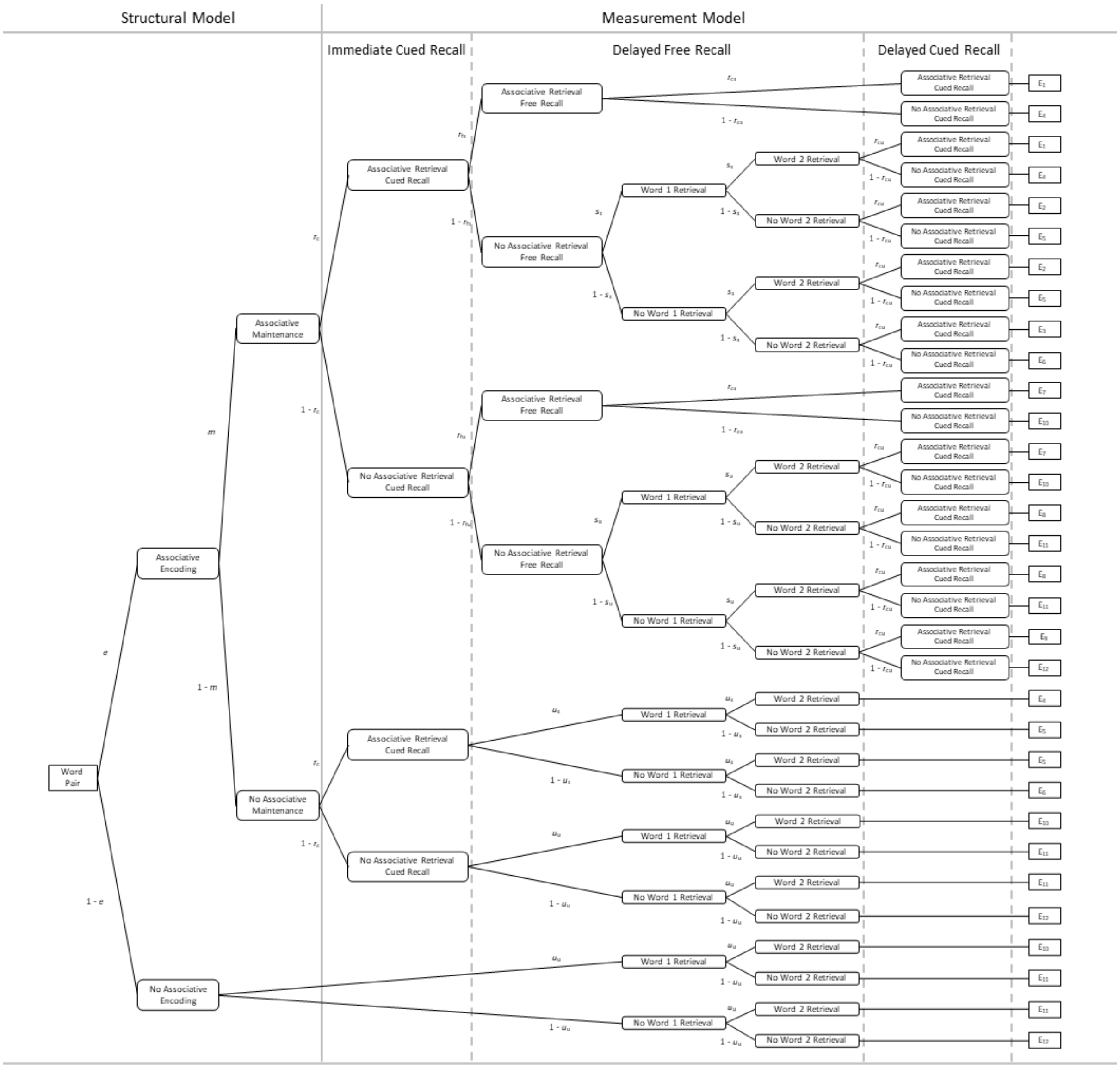
Note. Definition of the 12 observation categories E_1 to E_{12} adapted from “Encoding, Maintenance, and Retrieval Processes in the Lag Effect: A Multinomial Processing Tree Analysis”, by C. E. Küpper-Tetzel and E. Erdfelder, 2012, *Memory, 20*(1), p. 39 (https://doi.org/10.1080/09658211.2011.631550). Copyright 2012 by Psychology Press, Taylor & Francis.

In the EMR model, the underlying cognitive processes (i.e., encoding, maintenance, cued- and free recall retrieval) are conceived as binary, in other words, as either successful or unsuccessful. The model thus enumerates all possible combinations of encoding, maintenance, and retrieval outcomes that can occur in the extended free-then-cued-recall paradigm for each presented word pair. Based on this, it is possible to set up a system of model equations, expressing the probabilities of all 12 observation categories as a function of latent probabilities associated with the possible cognitive processes involved (i.e., encoding, maintenance, cued- and free recall retrieval).

As illustrated in Figure A1 (based on Erdfelder et al., 2022), the latent cognitive processes connect the stimulus (word pair) on the left with the observed participant response on the right. The generalized model version in Figure A1 also shows that the EMR model consists of a structural model (left side) and a measurement model (right side). The structural model defines three latent memory processes underlying memory judgments: First, a word-pair association has been successfully encoded (probability e) before immediate cued recall and is maintained in memory until delayed free-then-cued recall (probability $r_c \cdot m$). Second, a word-pair association has been successfully encoded (probability e) before immediate cued recall but is forgotten until delayed free-then-cued recall (probability $r_c \cdot (1-m)$). Third, a word-pair association has not been successfully encoded (complementary probability $1-e$) and is thus not available in memory for neither immediate cued recall nor delayed free-then-cued recall. By contrast, the measurement model defines the success probabilities in the immediate cued and delayed free-then-cued-recall tests. On the basis of previous research, Erdfelder et al. (2022) recommend the following procedure to specify the measurement model: First, the retrieval probability in the immediate cued recall test (probability r_c) depends only on successful encoding and not on processes that occur after the initial cued recall test. If encoding fails, immediate cued recall will be unsuccessful because word pairs are selected so that the second word cannot be guessed correctly when the word pair association is not available in memory. Second, an independence model in which the three latent states of encoding, maintenance, and retrieval fully explain stochastic dependencies between the immediate and the two delayed memory tests (cf. Küpper-Tetzel & Erdfelder, 2012) should be fitted first. This parsimonious model version includes 6 parameters only (i.e., parameter e, m, r_c, r_f, s, u), provides more reliable parameter estimates with smaller standard errors, and allows easier interpretation. However, in case the independence model produces misfit, the independence assumption should be relaxed, starting with parameter u and followed by parameter s. This procedure allows to cope with the possibility that word-pair storage fails in the encoding or maintenance phase whereas storage and retrieval of a single word in free recall may succeed depending on whether the immediate cued recall was successful (s_s, u_s) or unsuccessful (s_u, u_u). Moreover, it is assumed that delayed cued recall of an encoded association fails when a word-pair association is not successfully maintained during the retention interval. Third, because cued recall parameters (r_c) are expected to be close to 1 (i.e., cued recall retrieval of a stored word pair is typically successful), the restriction pattern $r_c = r_{cs} = r_{cu}$ is recommended as a default. Finally, it is important that immediate cued recall performance is neither too bad nor perfect as it causes zero cells in certain observation categories thereby threatening model applicability. For this reason, it is recommended to exclude participants with less than 35% correct responses in immediate cued recall to ensure sufficient observations in categories E_1 to E_6. Furthermore, participants with more than than 95% correct immediate cued recall responses should be excluded to reduce the occurrence of zero cells in categories E_7 to E_{12} (Erdfelder et al., 2022).

Figure A1

Generalized Version of the Encoding-Maintenance-Retrieval (EMR) Model for the Extended Free-Then-Cued Recall Paradigm



Note. Rectangles indicate observable test conditions and outcomes. Rectangles with rounded corners indicate latent cognitive processes. Parameters attached to the branches denote transition probabilities between unobservable states. Please refer to Table A1 for the meaning of the 12 possible outcome categories E_1 to E_{12}. *e* = probability of successful encoding of word-pair associations; *m* = probability of maintaining encoded associations across the retention interval; r_c = probability of retrieving stored associations in immediate cued recall; r_{cs} = probability of retrieving stored associations in delayed cued recall for successful immediate cued-recall retrieval; r_{cu} = probability of retrieving stored associations in delayed cued recall for unsuccessful immediate cued-recall retrieval; r_{fs} = probability of retrieving stored associations in free recall for successful immediate cued-recall retrieval; r_{fu} = probability of retrieving stored associations in free recall for unsuccessful immediate cued-recall retrieval; s_s = probability of associated single word retrieval during free recall for successful immediate cued-recall retrieval; s_u = probability of associated single word retrieval during free recall for unsuccessful immediate cued-recall retrieval; u_s = probability of non-associated single word retrieval during free recall for successful immediate cued-recall retrieval; u_u = probability of non-associated single word retrieval during free recall for unsuccessful immediate cued-recall retrieval. Adapted from “Why Does Sleep Improve Episodic Memory? An Encoding-Maintenance-Retrieval Analysis”, by E. Erdfelder, S. Berres, J. Quevedo Pütter, and C. E. Küpper-Tetzel, 2022, *Manuscript submitted for publication*.

Appendix B

Results of the Bayesian-Hierarchical Multinomial Processing Tree (MPT) Model Analyses

Table B1

Posterior Group Level Means and Posterior Standard Deviations of the Hierarchical Bayesian Multinomial Processing Tree (MPT) Analysis with TreeBUGS for Experiment 1

Parameter	Medium encoding		Strong encoding	
	Wake	Sleep	Wake	Sleep
e	**.61 (.06)**	**.58 (.06)**	**.90 (.04)**	**.91 (.03)**
m	**.91 (.03)**	**.96 (.09)**	**.95 (.03)**	**.94 (.14)**
r_c	.98 (.02)	.98 (.02)	.98 (.01)	.98 (.01)
r_f	**.23 (.04)**	**.36 (.05)**	**.42 (.05)**	**.42 (.07)**
s	.05 (.02)	.05 (.04)	.09 (.02)	.03 (.04)
u_s	.05 (.05)	.31 (.21)	.21 (.08)	.42 (.25)
u_u	.03 (.01)	.06 (.02)	.03 (.04)	.10 (.09)

Note. The table shows the estimated success probabilities of latent cognitive processes. Posterior standard deviations are presented in parentheses. Estimation results for core parameters are displayed in bold. There was good MCMC chain convergence ($\hat{R} < 1.05$) and model fit (medium-encoding-wake condition: $p_{T1} = .18$, $p_{T2} = .33$; medium-encoding-sleep condition: $p_{T1} = .37$, $p_{T2} = .64$; strong-encoding-wake condition: $p_{T1} = .53$, $p_{T2} = .61$; strong-encoding-sleep condition: $p_{T1} = .60$, $p_{T2} = .50$). e = probability of successful encoding of word-pair associations; m = probability of maintaining encoded associations across the retention interval; r_c = probability of retrieving stored associations in cued recall; r_f = probability of retrieving stored associations in free recall; s = probability of associated single word retrieval during free recall; u_s = probability of non-associated single word retrieval during free recall for successful immediate cued-recall retrieval; u_u = probability of non-associated single word retrieval during free recall for unsuccessful immediate cued-recall retrieval; $\hat{R}$

= potential scale reduction factor (Gelman & Rubin, 1992); p_{T1} = posterior predictive *p*-value for the mean; p_{T2} = posterior predictive *p*-value for the covariance.

Table B2

Posterior Group Level Means and Posterior Standard Deviations of the Hierarchical Bayesian Multinomial Processing Tree (MPT) Analysis with TreeBUGS for Experiment 2

Parameter	No test expectation		Test expectation	
	Wake	Sleep	Wake	Sleep
e	**.78 (.03)**	**.80 (.03)**	**.83 (.03)**	**.81 (.02)**
m	**.86 (.04)**	**.98 (.05)**	**.88 (.03)**	**.98 (.01)**
r_c	.99 (.01)	.98 (.01)	.99 (.01)	.99 (.01)
$\boldsymbol{r_f}$	**.52 (.03)**	**.52 (.05)**	**.47 (.03)**	**.50 (.02)**
s_s	.10 (.02)	.07 (.02)	.10 (.02)	.07 (.01)
s_u	.47 (.23)	.48 (.20)	.39 (.26)	.24 (.24)
u_s	.17 (.06)	.28 (.18)	.24 (.05)	.15 (.10)
u_u	.09 (.04)	.14 (.03)	.12 (.03)	.13 (.02)

Note. The table shows the estimated success probabilities of latent cognitive processes. Posterior standard deviations are presented in parentheses. Estimation results for core parameters are displayed in bold. There was good MCMC chain convergence ($\hat{R} < 1.05$) and model fit (no-expectation-wake condition: $p_{T1} = .44$, $p_{T2} = .47$; no-expectation-sleep condition: $p_{T1} = .24$, $p_{T2} = .47$; expectation-wake condition: $p_{T1} = .36$, $p_{T2} = .46$; expectation-sleep condition: $p_{T1} = .14$, $p_{T2} = .23$). e = probability of successful encoding of word-pair associations; m = probability of maintaining encoded associations across the retention interval; r_c = probability of retrieving stored associations in cued recall; r_f = probability of retrieving stored associations in free recall; s_s = probability of associated single word retrieval during free recall for successful immediate cued-recall retrieval; s_u = probability of associated single word retrieval during free recall for unsuccessful immediate cued-recall retrieval; u_s = probability of non-associated single word retrieval during free recall for successful immediate cued-recall retrieval; u_u = probability of non-associated single word retrieval during free recall for unsuccessful immediate cued-recall retrieval; $\hat{R}$ = potential scale reduction factor (Gelman

& Rubin, 1992); p_{T1} = posterior predictive *p*-value for the mean; p_{T2} = posterior predictive *p*-value for the covariance.

Supplemental Materials

accompanying the manuscript *Is it all About Storage? Effects of Encoding Strength and Test Expectation on the Sleep Benefit in Episodic Memory*

Experiment 1

Experiment 2

Table S1

Sample Characteristics for Experiment 1

Variable	Total		Wake		Sleep			
	M	*SD*	*M*	*SD*	*M*	*SD*	*t*(25)	*p*
Age, in years	22.89	6.46	24.40	8.40	21.00	1.35	1.38	.179
SSS, learning session	2.78	1.05	2.73	0.88	2.83	1.27	0.24	.811
SSS, testing session	2.81	1.00	2.80	1.08	2.83	0.94	0.08	.933
rMEQ, sum score	15.26	3.84	14.93	4.30	15.67	3.31	0.49	.631

Note. SSS = Stanford Sleepiness Scale (Hoddes et al., 1973); rMEQ = German version of the reduced Morningness-Eveningness Questionnaire (Randler, 2013). The SSS captures sleepiness on a 7-point scale with higher values indicating greater sleepiness. The rMEQ assesses chronotype with sum scores that are based on five items (Cronbach's α = .76). Sum scores can range from 4 to 25 with higher values indicating stronger morning preferences. The wake condition contains 8 female and 7 male participants, whereas the sleep condition contains only female participants.

Table S2

Sample Characteristics for Experiment 1 Including Only Participants Between 18 and 35 Years of Age

Variable	Total		Wake		Sleep			
	M	*SD*	*M*	*SD*	*M*	*SD*	*t*(23)	*p*
Age, in years	22.00	3.16	22.92	4.05	21.00	1.35	1.56	.131
SSS, learning session	2.76	1.05	2.69	0.85	2.83	1.27	0.33	.745
SSS, testing session	2.76	1.01	2.69	1.11	2.83	0.94	0.34	.736
rMEQ, sum score	15.44	3.84	15.23	4.40	15.67	3.31	0.28	.784
Sleep duration, self-report					7.54	0.61		
Sleep duration [a]					7.84	0.67		
Time awake [a]					0.01	0.02		
Deep sleep [a]					2.04	1.15		
Light sleep [a]					5.79	1.29		
Sleep quality					3.17	0.39		

Note. SSS = Stanford Sleepiness Scale (Hoddes et al., 1973); rMEQ = German version of the reduced Morningness-Eveningness Questionnaire (Randler, 2013). The SSS captures sleepiness on a 7-point scale with higher values indicating greater sleepiness. The rMEQ assesses chronotype with sum scores that are based on five items (Cronbach's α = .78). Sum scores can range from 4 to 25 with higher values indicating stronger morning preferences. Sleep duration, time awake, deep sleep, and light sleep are reported in hours. Sleep quality was assessed via self-report on a 5-point scale with higher values indicating better sleep quality. The wake condition contains 13 participants (8 female) and the sleep condition 12 participants (12 female). All participants stated that they slept for at least 6 hr in the night before the first session of the experiment and that they did not consume any alcoholic drinks between the first and second session. Two participants in the wake condition napped between the sessions: One participant napped for 28 min, the other for 1 hr.

[a] Assessed by a commercial fitness tracker (Xiaomi Mi Band 2).

Table S3

Means and Standard Errors for the Memory Performance Measures in Experiment 1 Including Only Participants Between 18 and 35 Years of Age

Dependent variable	Medium encoding		Strong encoding			
	Wake	Sleep	Wake	Sleep	*F*(1, 23)	*p*
Immediate cued recall	24.00 (1.42)	22.33 (1.32)	34.31 (0.92)	34.83 (0.68)	0.93	.344
Delayed cued recall	21.77 (1.48)	21.75 (1.37)	32.15 (1.17)	34.58 (0.87)	1.59	.221
Cued recall forgetting	2.23 (0.30)	0.58 (0.30)	2.15 (0.63)	0.25 (0.37)	0.07	.797
Free recall	6.15 (0.90)	8.08 (0.72)	15.15 (1.43)	14.75 (0.84)	1.21	.282

Note. The table shows the mean number of words (cued recall) and word pairs (free recall) correctly recalled in the respective memory test. Standard errors of the mean are presented in parentheses. Cued recall forgetting represents the difference between immediate and delayed cued recall performance. Participants were randomly assigned to a wake (*n* = 13) and sleep (*n* = 12) condition. *F* = *F* value for the interaction of encoding strength and wake versus sleep.

Table S4

Maximum-Likelihood Parameter Estimates and Standard Errors of the Encoding-Maintenance-Retrieval (EMR) Multinomial Model for Experiment 1 Including Only Participants Between 18 and 35 Years of Age

Parameter	Medium encoding		Strong encoding	
	Wake	Sleep	Wake	Sleep
e	.62 (.02)	.57 (.02)	.88 (.01)	.89 (.01)
m	.90 (.02)	.97 (.01)	.93 (.01)	.99 (.01)
r_c	.98 (.01)	.98 (.01)	.98 (.01)	.99 (.01)
r_f	.27 (.03)	.36 (.03)	.46 (.02)	.42 (.02)
s	.06 (.01)	.04 (.01)	.08 (.01)	.03 (.01)
u_s	.06 (.03)	.22 (.11)	.20 (.06)	.23 (.19)
u_u	.03 (.01)	.06 (.01)	.03 (.01)	.09 (.03)

Note. The table shows the estimated success probabilities of latent cognitive processes. Standard errors of the parameter estimates are presented in parentheses. The model fitted the data well, $G^2(16) = 20.86$, $p = .184$. e = probability of successful encoding of word-pair associations; m = probability of maintaining encoded associations across the retention interval; r_c = probability of retrieving stored associations in cued recall; r_f = probability of retrieving stored associations in free recall; s = probability of associated single word retrieval during free recall; u_s = probability of non-associated single word retrieval during free recall for successful immediate cued-recall retrieval; u_u = probability of non-associated single word retrieval during free recall for unsuccessful immediate cued-recall retrieval.

Table S5

Posterior Group Level Means and Posterior Standard Deviations of the Hierarchical Bayesian Multinomial Processing Tree (MPT) Analysis with TreeBUGS for Experiment 1 Including Only Participants Between 18 and 35 Years of Age

Parameter	Medium encoding		Strong encoding	
	Wake	Sleep	Wake	Sleep
e	.64 (.07)	.58 (.05)	.90 (.04)	.91 (.03)
m	.92 (.04)	.96 (.11)	.96 (.04)	.95 (.13)
r_c	.98 (.03)	.98 (.02)	.98 (.01)	.99 (.01)
r_f	.23 (.05)	.36 (.05)	.44 (.06)	.42 (.06)
s	.06 (.02)	.05 (.05)	.09 (.02)	.03 (.03)
u_s	.08 (.09)	.31 (.21)	.22 (.09)	.43 (.26)
u_u	.03 (.02)	.06 (.03)	.04 (.06)	.10 (.09)

Note. The table shows the estimated success probabilities of latent cognitive processes. Posterior standard deviations are presented in parentheses. There was good MCMC chain convergence ($\hat{R} < 1.05$) and model fit (medium-encoding-wake condition: $p_{T1} = .27$, $p_{T2} = .37$; medium-encoding-sleep condition: $p_{T1} = .35$, $p_{T2} = .64$; strong-encoding-wake condition: $p_{T1} = .42$, $p_{T2} = .59$; strong-encoding-sleep condition: $p_{T1} = .60$, $p_{T2} = .50$). e = probability of successful encoding of word-pair associations; m = probability of maintaining encoded associations across the retention interval; r_c = probability of retrieving stored associations in cued recall; r_f = probability of retrieving stored associations in free recall; s = probability of associated single word retrieval during free recall; u_s = probability of non-associated single word retrieval during free recall for successful immediate cued-recall retrieval; u_u = probability of non-associated single word retrieval during free recall for unsuccessful immediate cued-recall retrieval; $\hat{R}$ = potential scale reduction factor (Gelman & Rubin, 1992); p_{T1} = posterior predictive *p*-value for the mean; p_{T2} = posterior predictive *p*-value for the covariance.

Table S6

Sample Characteristics for Experiment 1 Including Only Participants Without Daytime Naps

Variable	Total		Wake		Sleep			
	M	*SD*	*M*	*SD*	*M*	*SD*	*t*(23)	*p*
Age, in years	22.88	6.72	24.62	9.04	21.00	1.35	1.37	.184
SSS, learning session	2.80	1.08	2.77	0.93	2.83	1.27	0.15	.886
SSS, testing session	2.92	0.95	3.00	1.00	2.83	0.94	0.43	.672
rMEQ, sum score	15.28	3.95	14.92	4.57	15.67	3.31	0.46	.648
Sleep duration, self-report					7.54	0.61		
Sleep duration [a]					7.84	0.67		
Time awake [a]					0.01	0.02		
Deep sleep [a]					2.04	1.15		
Light sleep [a]					5.79	1.29		
Sleep quality					3.17	0.39		

Note. SSS = Stanford Sleepiness Scale (Hoddes et al., 1973); rMEQ = German version of the reduced Morningness-Eveningness Questionnaire (Randler, 2013). The SSS captures sleepiness on a 7-point scale with higher values indicating greater sleepiness. The rMEQ assesses chronotype with sum scores that are based on five items (Cronbach's α = .76). Sum scores can range from 4 to 25 with higher values indicating stronger morning preferences. Sleep duration, time awake, deep sleep, and light sleep are reported in hours. Sleep quality was assessed via self-report on a 5-point scale with higher values indicating better sleep quality. The wake condition contains 13 participants (7 female) and the sleep condition 12 participants (12 female). All participants stated that they slept for at least 6 hr in the night before the first session of the experiment and that they did not consume any alcoholic drinks between the first and second session.

[a] Assessed by a commercial fitness tracker (Xiaomi Mi Band 2).

Table S7

Means and Standard Errors for the Memory Performance Measures in Experiment 1 Including Only Participants Without Daytime Naps

Dependent variable	Medium encoding		Strong encoding			
	Wake	Sleep	Wake	Sleep	*F*(1, 23)	*p*
Immediate cued recall	22.69 (1.53)	22.33 (1.32)	34.31 (0.97)	34.83 (0.68)	0.14	.714
Delayed cued recall	20.23 (1.57)	21.75 (1.37)	31.77 (1.16)	34.58 (0.87)	0.45	.507
Cued recall forgetting	2.46 (0.25)	0.58 (0.30)	2.54 (0.58)	0.25 (0.37)	0.18	.672
Free recall	5.62 (0.92)	8.08 (0.72)	14.77 (1.47)	14.75 (0.84)	1.40	.250

Note. The table shows the mean number of words (cued recall) and word pairs (free recall) correctly recalled in the respective memory test. Standard errors of the mean are presented in parentheses. Cued recall forgetting represents the difference between immediate and delayed cued recall performance. Participants were randomly assigned to a wake (*n* = 13) and sleep (*n* = 12) condition. *F* = *F* value for the interaction of encoding strength and wake versus sleep.

Table S8

Maximum-Likelihood Parameter Estimates and Standard Errors of the Encoding-Maintenance-Retrieval (EMR) Multinomial Model for Experiment 1 Including Only Participants Without Daytime Naps

Parameter	Medium encoding		Strong encoding	
	Wake	Sleep	Wake	Sleep
e	.58 (.02)	.57 (.02)	.88 (.01)	.89 (.01)
m	.88 (.02)	.97 (.01)	.92 (.01)	.99 (.01)
r_c	.98 (.01)	.98 (.01)	.98 (.01)	.99 (.01)
r_f	.27 (.03)	.36 (.03)	.45 (.02)	.42 (.02)
s	.06 (.01)	.04 (.01)	.09 (.01)	.03 (.01)
u_s	.05 (.03)	.22 (.11)	.20 (.06)	.23 (.19)
u_u	.03 (.01)	.06 (.01)	.03 (.02)	.09 (.03)

Note. The table shows the estimated success probabilities of latent cognitive processes. Standard errors of the parameter estimates are presented in parentheses. The model fitted the data well, $G^2(16) = 18.54$, $p = .293$. e = probability of successful encoding of word-pair associations; m = probability of maintaining encoded associations across the retention interval; r_c = probability of retrieving stored associations in cued recall; r_f = probability of retrieving stored associations in free recall; s = probability of associated single word retrieval during free recall; u_s = probability of non-associated single word retrieval during free recall for successful immediate cued-recall retrieval; u_u = probability of non-associated single word retrieval during free recall for unsuccessful immediate cued-recall retrieval.

Table S9

Posterior Group Level Means and Posterior Standard Deviations of the Hierarchical Bayesian Multinomial Processing Tree (MPT) Analysis with TreeBUGS for Experiment 1 Including Only Participants Without Daytime Naps

Parameter	Medium encoding		Strong encoding	
	Wake	Sleep	Wake	Sleep
e	.61 (.07)	.58 (.06)	.90 (.05)	.91 (.03)
m	.90 (.03)	.96 (.09)	.95 (.03)	.95 (.13)
r_c	.98 (.03)	.98 (.02)	.98 (.01)	.99 (.01)
r_f	.23 (.05)	.36 (.05)	.42 (.07)	.42 (.06)
s	.05 (.02)	.05 (.04)	.10 (.02)	.03 (.04)
u_s	.06 (.06)	.30 (.21)	.24 (.09)	.42 (.26)
u_u	.03 (.02)	.06 (.03)	.05 (.06)	.10 (.09)

Note. The table shows the estimated success probabilities of latent cognitive processes. Posterior standard deviations are presented in parentheses. There was good MCMC chain convergence ($\hat{R} < 1.05$) and model fit (medium-encoding-wake condition: $p_{T1} = .26$, $p_{T2} = .35$; medium-encoding-sleep condition: $p_{T1} = .35$, $p_{T2} = .64$; strong-encoding-wake condition: $p_{T1} = .56$, $p_{T2} = .61$; strong-encoding-sleep condition: $p_{T1} = .60$, $p_{T2} = .49$). e = probability of successful encoding of word-pair associations; m = probability of maintaining encoded associations across the retention interval; r_c = probability of retrieving stored associations in cued recall; r_f = probability of retrieving stored associations in free recall; s = probability of associated single word retrieval during free recall; u_s = probability of non-associated single word retrieval during free recall for successful immediate cued-recall retrieval; u_u = probability of non-associated single word retrieval during free recall for unsuccessful immediate cued-recall retrieval; $\hat{R}$ = potential scale reduction factor (Gelman & Rubin, 1992); p_{T1} = posterior predictive *p*-value for the mean; p_{T2} = posterior predictive *p*-value for the covariance.

Table S10

Sample Characteristics for Experiment 1 Including all Participants Previously Excluded due EMR Model Specific Exclusion Criteria

Variable	Total		Wake		Sleep			
	M	*SD*	*M*	*SD*	*M*	*SD*	*t*(65)	*p*
Age, in years	22.87	5.69	24.22	7.75	21.63	2.17	1.90	.062
SSS, learning session	2.84	0.93	2.81	0.93	2.86	0.94	0.19	.846
SSS, testing session	2.73	1.12	3.03	1.23	2.46	0.95	2.15	.035
rMEQ, sum score	14.88	4.05	14.75	3.88	15.00	4.26	0.25	.803
Sleep duration, self-report					7.24	0.89		
Sleep duration [a]					7.55	0.90		
Time awake [a]					0.04	0.13		
Deep sleep [a]					2.12	1.12		
Light sleep [a]					5.43	1.27		
Sleep quality					3.29	0.86		

Note. SSS = Stanford Sleepiness Scale (Hoddes et al., 1973); rMEQ = German version of the reduced Morningness-Eveningness Questionnaire (Randler, 2013). The SSS captures sleepiness on a 7-point scale with higher values indicating greater sleepiness. The rMEQ assesses chronotype with sum scores that are based on five items (Cronbach's α = .78). Sum scores can range from 4 to 25 with higher values indicating stronger morning preferences. Sleep duration, time awake, deep sleep, and light sleep are reported in hours. Sleep quality was assessed via self-report on a 5-point scale with higher values indicating better sleep quality. The wake condition contains 32 participants (18 female) and the sleep condition 35 participants (30 female). All participants stated that they slept for at least 6 hr in the night before the first session of the experiment and that they did not consume any alcoholic drinks between the first and second session.

[a] Assessed by a commercial fitness tracker (Xiaomi Mi Band 2).

Table S11

Means and Standard Errors for the Memory Performance Measures in Experiment 1 Including all Participants Previously Excluded due EMR Model Specific Exclusion Criteria

Dependent variable	Medium encoding		Strong encoding			
	Wake	Sleep	Wake	Sleep	*F*(1, 65)	*p*
Immediate cued recall	21.56 (1.33)	24.86 (1.24)	32.72 (0.99)	35.40 (0.82)	0.16	.694
Delayed cued recall	19.34 (1.31)	24.29 (1.26)	30.06 (1.11)	34.71 (0.92)	0.04	.833
Cued recall forgetting	2.22 (0.23)	0.57 (0.17)	2.66 (0.35)	0.69 (0.21)	0.34	.560
Free recall	5.28 (0.46)	8.43 (0.56)	12.59 (0.78)	15.09 (0.65)	0.32	.574

Note. The table shows the mean number of words (cued recall) and word pairs (free recall) correctly recalled in the respective memory test. Standard errors of the mean are presented in parentheses. Cued recall forgetting represents the difference between immediate and delayed cued recall performance. Participants were randomly assigned to a wake (*n* = 32) and sleep (*n* = 35) condition. *F* = *F* value for the interaction of encoding strength and wake versus sleep.

Additional Memory Performance Analysis for Experiment 1 Including all Participants Previously Excluded due EMR Model Specific Exclusion Criteria

Due to the application of the exclusion criteria recommended for the EMR model, 40 participants had to be excluded. We therefore performed all ANOVAs for the memory performance analysis again, this time including the 40 previously excluded participants (i.e., new sample size: $n_{\text{sleep}} = 35$, $n_{\text{wake}} = 32$). As in the main analysis, a statistically significant main effect of encoding strength emerged with better memory performance in strong- compared to medium encoding, delayed cued recall: $F(1, 65) = 237.84$, $MSE = 15.72$, $p < .001$, $\eta_p^2 = .79$; free recall: $F(1, 65) = 145.16$, $MSE = 11.24$, $p < .001$, $\eta_p^2 = .69$. In contrast to the main analysis, we found a statistically significant main effect of wake versus sleep with better memory performance after sleep compared to wakefulness, delayed cued recall: $F(1, 65) = 4.70$, $MSE = 163.57$, $p = .034$, $\eta_p^2 = .07$; free recall: $F(1, 65) = 6.49$, $MSE = 40.94$, $p = .013$, $\eta_p^2 = .09$. In line with the results of the main analysis, however, no statistically significant interaction effects of encoding strength and wake versus sleep emerged; delayed cued recall: $F(1, 65) = 0.04$, $MSE = 15.72$, $p = .833$, $\eta_p^2 < .01$; free recall: $F(1, 65) = 0.32$, $MSE = 11.24$, $p = .574$, $\eta_p^2 = .01$. Furthermore, the result patterns for cued recall forgetting did not differ between the main and current analyses; main effect of encoding strength: $F(1, 65) = 1.00$, $MSE = 2.55$, $p = .321$, $\eta_p^2 = .02$, main effect of wake versus sleep: $F(1, 65) = 19.58$, $MSE = 5.59$, $p < .001$, $\eta_p^2 = .23$, interaction effect of encoding strength and wake versus sleep: $F(1, 65) = 0.34$, $MSE = 2.55$, $p = .560$, $\eta_p^2 = .01$. The same was true for immediate cued recall; main effect of encoding strength: $F(1, 65) = 196.10$, $MSE = 20.07$, $p < .001$, $\eta_p^2 = .75$, main effect of wake versus sleep: $F(1, 65) = 2.06$, $MSE = 144.62$, $p = .156$, $\eta_p^2 = .03$, interaction effect of encoding strength and wake versus sleep: $F(1, 65) = 0.16$, $MSE = 20.07$, $p = .694$, $\eta_p^2 = .002$. Thus, similar results emerged when using the larger sample size of $N = 67$ (for sample characteristics, sample means and standard errors of all experimental conditions, see Tables S10 and S11).

Table S12

Sample Characteristics for Experiment 2

Variable	Total		No test expectation				Test expectation					
			Wake		Sleep		Wake		Sleep			
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*(3, 66)	*p*
Age, in years	22.06	4.81	23.53	8.29	22.31	3.07	21.89	3.31	20.91	2.51	0.97	.414
SSS, learning session	2.79	1.06	2.41	1.06	2.92	1.32	2.61	0.92	3.14	0.94	1.80	.156
SSS, testing session	2.77	1.22	3.18	0.81	2.54	1.39	3.11	1.41	2.32	1.09	2.41	.075
rMEQ, sum score	15.03	3.58	14.88	4.06	14.62	4.17	14.22	3.28	16.05	3.03	0.96	.417

Note. SSS = Stanford Sleepiness Scale (Hoddes et al., 1973); rMEQ = German version of the reduced Morningness-Eveningness Questionnaire (Randler, 2013). The SSS captures sleepiness on a 7-point scale with higher values indicating greater sleepiness. The rMEQ assesses chronotype with sum scores that are based on five items (Cronbach's α = .71). Sum scores can range from 4 to 25 with higher values indicating stronger morning preferences. The no-expectation-wake condition contains 14 female and 3 male participants, the no-expectation-sleep condition 8 female and 5 male participants, the expectation-wake condition 13 female and 5 male participants, and the expectation-sleep condition 16 female and 6 male participants.

Table S13

Sample Characteristics for Experiment 2 Including Only Participants Between 18 and 35 Years of Age

Variable	Total		No test expectation				Test expectation						
			Wake		Sleep		Wake		Sleep				
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*(3, 65)	*t*(33)	*p*
Age, in years	21.59	2.88	21.62	2.75	22.31	3.07	21.89	3.31	20.91	2.51	0.74		.533
SSS, learning session	2.78	1.07	2.38	1.09	2.92	1.32	2.61	0.92	3.14	0.94	1.88		.142
SSS, testing session	2.75	1.22	3.12	0.81	2.54	1.39	3.11	1.41	2.32	1.09	2.20		.097
rMEQ, sum score	14.96	3.56	14.56	3.97	14.62	4.17	14.22	3.28	16.05	3.03	1.05		.377
Sleep duration, self-report					7.42	0.81			7.48	0.65		0.22	.829
Sleep duration [a]					7.82	0.94			7.88	0.74		0.19[c]	.847
Time awake [a]					0.01	0.02			0.04	0.14		0.74[c]	.465
Deep sleep [a]					1.94	1.15			2.14	0.95		0.55[c]	.588
Light sleep [a]					5.88	1.15			5.74	0.92		0.40[c]	.691
Sleep quality					3.42	1.00			3.55	0.96		0.37[c]	.715
Repetitions [b]	0.68	0.88	0.56	0.81	0.46	0.66	1.00	0.77	0.64	1.09	1.18		.326

Note. SSS = Stanford Sleepiness Scale (Hoddes et al., 1973); rMEQ = German version of the reduced Morningness-Eveningness Questionnaire (Randler, 2013). The SSS captures sleepiness on a 7-point scale with higher values indicating greater sleepiness. The rMEQ assesses chronotype with sum scores that are based on five items (Cronbach's $\alpha = .70$). Sum scores can range from 4 to 25 with higher values indicating stronger morning preferences. Sleep duration, time awake, deep sleep, and light sleep are reported in hours. Sleep quality was assessed via self-report on a 5-point scale with higher values indicating better sleep quality. The no-expectation-wake condition contains 16 participants (13 female), the no-expectation-sleep condition 13 participants (8 female), the expectation-wake condition 18 participants (13 female), and the expectation-sleep condition 22 participants (16 female). With the exception of 8 participants ($n_{\text{no expectation, wake}} = 1$, $n_{\text{no expectation, sleep}} = 3$, $n_{\text{expectation, wake}} = 3$, $n_{\text{expectation, sleep}} = 1$), all participants stated that they slept for at least 6 hr in the night before the first session of the experiment. All but one participant reported that they did not consume any alcoholic drinks between the first and second session. Two participants in the wake condition napped between the sessions: One participant napped for 30 min, the other for 1 hr.

[a] Assessed by a commercial fitness tracker (Xiaomi Mi Band 2). [b] How many times the sequence of learning phase, distractor task, and immediate cued recall task was repeated until participants responded at least 60% correctly on the immediate cued recall task. [c] $df = 32$, because sleep quality was not captured for one participant and the fitness tracker did not record any data for another participant.

Table S14

Means and Standard Errors for the Memory Performance Measures in Experiment 2 Including Only Participants Between 18 and 35 Years of Age

Dependent variable	No test expectation		Test expectation			
	Wake (*n* = 16)	Sleep (*n* = 13)	Wake (*n* = 18)	Sleep (*n* = 22)	*F*(1, 65)	*p*
Immediate cued recall	30.62 (0.48)	30.54 (0.51)	32.39 (0.57)	31.50 (0.40)	0.17	.685
Delayed cued recall	25.94 (0.74)	29.54 (0.58)	28.00 (0.80)	30.32 (0.52)	0.22	.638
Cued recall forgetting	4.69 (0.44)	1.00 (0.26)	4.39 (0.31)	1.18 (0.21)	0.14	.705
Free recall	14.19 (0.64)	16.08 (0.67)	13.94 (0.65)	15.32 (0.43)	0.05	.830

Note. The table shows the mean number of words (cued recall) and word pairs (free recall) correctly recalled in the respective memory test. Standard errors of the mean are presented in parentheses. Cued recall forgetting represents the difference between immediate and delayed cued recall performance. *F* = *F* value for the interaction of test expectation and wake versus sleep.

Table S15

Maximum-Likelihood Parameter Estimates and Standard Errors of the Encoding-Maintenance-Retrieval (EMR) Multinomial Model for Experiment 2 Including Only Participants Between 18 and 35 Years of Age

Parameter	No test expectation		Test expectation	
	Wake	Sleep	Wake	Sleep
e	.77 (.02)	.78 (.02)	.82 (.01)	.80 (.01)
m	.85 (.02)	.96 (.01)	.87 (.01)	.96 (.01)
r_c	.99 (.01)	.98 (.01)	.99 (.01)	.99 (.01)
r_f	.53 (.02)	.53 (.03)	.48 (.02)	.50 (.02)
s_s	.09 (.02)	.06 (.01)	.10 (.01)	.06 (.01)
s_u	.60 (.25)	.51 (.19)	.20 (.19)	.00 (.49)
u_s	.15 (.03)	.16 (.08)	.23 (.04)	.15 (.05)
u_u	.10 (.02)	.13 (.02)	.11 (.02)	.13 (.02)

Note. The table shows the estimated success probabilities of latent cognitive processes. Standard errors of the parameter estimates are presented in parentheses. The model produced a slight misfit, $G^2(12) = 21.71$, $p = .041$. e = probability of successful encoding of word-pair associations; m = probability of maintaining encoded associations across the retention interval; r_c = probability of retrieving stored associations in cued recall; r_f = probability of retrieving stored associations in free recall; s_s = probability of associated single word retrieval during free recall for successful immediate cued-recall retrieval; s_u = probability of associated single word retrieval during free recall for unsuccessful immediate cued-recall retrieval; u_s = probability of non-associated single word retrieval during free recall for successful immediate cued-recall retrieval; u_u = probability of non-associated single word retrieval during free recall for unsuccessful immediate cued-recall retrieval.

Table S16

Posterior Group Level Means and Posterior Standard Deviations of the Hierarchical Bayesian Multinomial Processing Tree (MPT) Analysis with TreeBUGS for Experiment 2 Including Only Participants Between 18 and 35 Years of Age

Parameter	No test expectation		Test expectation	
	Wake	Sleep	Wake	Sleep
e	.79 (.02)	.80 (.03)	.83 (.03)	.81 (.02)
m	.87 (.04)	.98 (.04)	.88 (.03)	.98 (.02)
r_c	.99 (.01)	.98 (.01)	.99 (.01)	.99 (.01)
r_f	.52 (.04)	.52 (.05)	.47 (.03)	.50 (.02)
s_s	.10 (.02)	.07 (.02)	.10 (.02)	.07 (.01)
s_u	.56 (.23)	.49 (.21)	.38 (.25)	.23 (.24)
u_s	.17 (.06)	.28 (.18)	.24 (.05)	.15 (.10)
u_u	.10 (.04)	.14 (.03)	.12 (.03)	.13 (.02)

Note. The table shows the estimated success probabilities of latent cognitive processes. Posterior standard deviations are presented in parentheses. There was good MCMC chain convergence ($\hat{R}$ < 1.05) and model fit (no-expectation-wake condition: p_{T1} = .34, p_{T2} = .44; no-expectation-sleep condition: p_{T1} = .24, p_{T2} = .45; expectation-wake condition: p_{T1} = .37, p_{T2} = .47; expectation-sleep condition: p_{T1} = .14, p_{T2} = .23). e = probability of successful encoding of word-pair associations; m = probability of maintaining encoded associations across the retention interval; r_c = probability of retrieving stored associations in cued recall; r_f = probability of retrieving stored associations in free recall; s_s = probability of associated single word retrieval during free recall for successful immediate cued-recall retrieval; s_u = probability of associated single word retrieval during free recall for unsuccessful immediate cued-recall retrieval; u_s = probability of non-associated single word retrieval during free recall for successful immediate cued-recall retrieval; u_u = probability of non-associated single word retrieval during free recall for unsuccessful immediate cued-recall retrieval; $\hat{R}$ = potential scale reduction factor (Gelman & Rubin, 1992); p_{T1} = posterior predictive *p*-value for the mean; p_{T2} = posterior predictive *p*-value for the covariance.

Table S17

Sample Characteristics for Experiment 2 Including Only Participants Without Daytime Naps

Variable	Total		No test expectation				Test expectation						
			Wake		Sleep		Wake		Sleep				
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*(3, 64)	*t*(33)	*p*
Age, in years	22.04	4.86	23.67	8.80	22.31	3.07	21.89	3.31	20.91	2.51	0.97		.411
SSS, learning session	2.78	1.06	2.33	1.05	2.92	1.32	2.61	0.92	3.14	0.94	2.03		.119
SSS, testing session	2.78	1.23	3.27	0.80	2.54	1.39	3.11	1.41	2.32	1.09	2.58		.061
rMEQ, sum score	15.13	3.56	15.33	4.01	14.62	4.17	14.22	3.28	16.05	3.03	0.98		.407
Sleep duration, self-report					7.42	0.81			7.48	0.65		0.22	.829
Sleep duration [a]					7.82	0.94			7.88	0.74		0.19[c]	.847
Time awake [a]					0.01	0.02			0.04	0.14		0.74[c]	.465
Deep sleep [a]					1.94	1.15			2.14	0.95		0.55[c]	.588
Light sleep [a]					5.88	1.15			5.74	0.92		0.40[c]	.691
Sleep quality					3.42	1.00			3.55	0.96		0.37[c]	.715
Repetitions [b]	0.69	0.89	0.60	0.83	0.46	0.66	1.00	0.77	0.64	1.09	1.11		.352

Note. SSS = Stanford Sleepiness Scale (Hoddes et al., 1973); rMEQ = German version of the reduced Morningness-Eveningness Questionnaire (Randler, 2013). The SSS captures sleepiness on a 7-point scale with higher values indicating greater sleepiness. The rMEQ assesses chronotype with sum scores that are based on five items (Cronbach's α = .71). Sum scores can range from 4 to 25 with higher values indicating stronger morning preferences. Sleep duration, time awake, deep sleep, and light sleep are reported in hours. Sleep quality was assessed via self-report on a 5-point scale with higher values indicating better sleep quality. The no-expectation-wake condition contains 15 participants (14 female), the no-expectation-sleep condition 13 participants (8 female), the expectation-wake condition 18 participants (13 female), and the expectation-sleep condition 22 participants (16 female). With the exception of 8 participants ($n_{\text{no expectation, wake}} = 1$, $n_{\text{no expectation, sleep}} = 3$, $n_{\text{expectation, wake}} = 3$, $n_{\text{expectation, sleep}} = 1$), all participants stated that they slept for at least 6 hr in the night before the first session of the experiment. All but one participant reported that they did not consume any alcoholic drinks between the first and second session.

[a] Assessed by a commercial fitness tracker (Xiaomi Mi Band 2). [b] How many times the sequence of learning phase, distractor task, and immediate cued recall task was repeated until participants responded at least 60% correctly on the immediate cued recall task. [c] *df* = 32, because sleep quality was not captured for one participant and the fitness tracker did not record any data for another participant.

Table S18

Means and Standard Errors for the Memory Performance Measures in Experiment 2 Including Only Participants Without Daytime Naps

Dependent variable	No test expectation		Test expectation			
	Wake (*n* = 15)	Sleep (*n* = 13)	Wake (*n* = 18)	Sleep (*n* = 22)	*F*(1, 64)	*p*
Immediate cued recall	30.47 (0.51)	30.54 (0.51)	32.39 (0.57)	31.50 (0.40)	0.23	.636
Delayed cued recall	25.33 (0.79)	29.54 (0.58)	28.00 (0.81)	30.32 (0.52)	0.46	.499
Cued recall forgetting	5.13 (0.44)	1.00 (0.26)	4.39 (0.31)	1.18 (0.21)	0.54	.464
Free recall	14.00 (0.66)	16.08 (0.68)	13.94 (0.65)	15.32 (0.43)	0.08	.773

Note. The table shows the mean number of words (cued recall) and word pairs (free recall) correctly recalled in the respective memory test. Standard errors of the mean are presented in parentheses. Cued recall forgetting represents the difference between immediate and delayed cued recall performance. *F* = *F* value for the interaction of test expectation and wake versus sleep.

Table S19

Maximum-Likelihood Parameter Estimates and Standard Errors of the Encoding-Maintenance-Retrieval (EMR) Multinomial Model for Experiment 2 Including Only Participants Without Daytime Naps

Parameter	No test expectation		Test expectation	
	Wake	Sleep	Wake	Sleep
e	.77 (.02)	.78 (.02)	.82 (.01)	.80 (.01)
m	.84 (.02)	.96 (.01)	.87 (.01)	.96 (.01)
r_c	.98 (.01)	.98 (.01)	.99 (.01)	.99 (.01)
r_f	.53 (.03)	.53 (.03)	.48 (.02)	.50 (.02)
s_s	.10 (.02)	.06 (.01)	.10 (.01)	.06 (.01)
s_u	.41 (.22)	.51 (.19)	.20 (.19)	.00 (.49)
u_s	.16 (.03)	.16 (.08)	.23 (.04)	.15 (.05)
u_u	.10 (.02)	.13 (.02)	.11 (.02)	.13 (.02)

Note. The table shows the estimated success probabilities of latent cognitive processes. Standard errors of the parameter estimates are presented in parentheses. The model produced a slight misfit, $G^2(12) = 21.77$, $p = .040$. e = probability of successful encoding of word-pair associations; m = probability of maintaining encoded associations across the retention interval; r_c = probability of retrieving stored associations in cued recall; r_f = probability of retrieving stored associations in free recall; s_s = probability of associated single word retrieval during free recall for successful immediate cued-recall retrieval; s_u = probability of associated single word retrieval during free recall for unsuccessful immediate cued-recall retrieval; u_s = probability of non-associated single word retrieval during free recall for successful immediate cued-recall retrieval; u_u = probability of non-associated single word retrieval during free recall for unsuccessful immediate cued-recall retrieval.

Table S20

Posterior Group Level Means and Posterior Standard Deviations of the Hierarchical Bayesian Multinomial Processing Tree (MPT) Analysis with TreeBUGS for Experiment 2 Including Only Participants Without Daytime Naps

Parameter	No test expectation		Test expectation	
	Wake	Sleep	Wake	Sleep
e	.79 (.03)	.80 (.03)	.83 (.03)	.81 (.02)
m	.85 (.04)	.98 (.04)	.88 (.03)	.98 (.02)
r_c	.98 (.01)	.98 (.01)	.99 (.01)	.99 (.01)
r_f	.51 (.04)	.52 (.05)	.47 (.03)	.50 (.02)
s_s	.10 (.02)	.07 (.02)	.10 (.02)	.07 (.01)
s_u	.48 (.23)	.48 (.20)	.39 (.26)	.24 (.24)
u_s	.17 (.06)	.28 (.18)	.24 (.05)	.15 (.10)
u_u	.10 (.04)	.14 (.03)	.12 (.03)	.13 (.02)

Note. The table shows the estimated success probabilities of latent cognitive processes. Posterior standard deviations are presented in parentheses. There was good MCMC chain convergence ($\hat{R} <$ 1.05) and model fit (no-expectation-wake condition: $p_{T1} = .47$, $p_{T2} = .48$; no-expectation-sleep condition: $p_{T1} = .25$, $p_{T2} = .47$; expectation-wake condition: $p_{T1} = .37$, $p_{T2} = .46$; expectation-sleep condition: $p_{T1} = .13$, $p_{T2} = .24$). e = probability of successful encoding of word-pair associations; m = probability of maintaining encoded associations across the retention interval; r_c = probability of retrieving stored associations in cued recall; r_f = probability of retrieving stored associations in free recall; s_s = probability of associated single word retrieval during free recall for successful immediate cued-recall retrieval; s_u = probability of associated single word retrieval during free recall for unsuccessful immediate cued-recall retrieval; u_s = probability of non-associated single word retrieval during free recall for successful immediate cued-recall retrieval; u_u = probability of non-associated single word retrieval during free recall for unsuccessful immediate cued-recall retrieval; $\hat{R}$ = potential scale reduction factor (Gelman & Rubin, 1992); p_{T1} = posterior predictive *p*-value for the mean; p_{T2} = posterior predictive *p*-value for the covariance.

Table S21

Sample Characteristics for Experiment 2 Including all Participants Previously Excluded due EMR Model Specific Exclusion Criteria

Variable	Total		No test expectation				Test expectation						
			Wake		Sleep		Wake		Sleep				
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*(3, 70)	*t*(35)	*p*
Age, in years	22.12	4.70	23.61	8.05	22.29	2.95	21.84	3.22	21.09	2.59	1.00		.398
SSS, learning session	2.80	1.03	2.44	1.04	2.93	1.27	2.63	0.90	3.13	0.92	1.79		.157
SSS, testing session	2.74	1.25	3.28	0.89	2.43	1.40	3.00	1.45	2.30	1.06	2.80		.046
rMEQ, sum score	14.96	3.63	14.67	4.04	14.64	4.01	13.95	3.41	16.22	3.07	1.52		.218
Sleep duration, self-report					7.48	0.81			7.49	0.64		0.03	.977
Sleep duration [a]					7.80	0.90			7.87	0.72		0.24[c]	.812
Time awake [a]					0.01	0.02			0.04	0.14		0.76[c]	.451
Deep sleep [a]					1.99	1.12			2.18	0.95		0.53[c]	.597
Light sleep [a]					5.81	1.13			5.69	0.93		0.35[c]	.730
Sleep quality					3.46	0.97			3.57	0.95		0.31[c]	.756
Repetitions [b]	0.70	0.86	0.61	0.78	0.50	0.65	1.00	0.75	0.65	1.07	1.13		.344

Note. SSS = Stanford Sleepiness Scale (Hoddes et al., 1973); rMEQ = German version of the reduced Morningness-Eveningness Questionnaire (Randler, 2013). The SSS captures sleepiness on a 7-point scale with higher values indicating greater sleepiness. The rMEQ assesses chronotype with sum scores that are based on five items (Cronbach's α = .73). Sum scores can range from 4 to 25 with higher values indicating stronger morning preferences. Sleep duration, time awake, deep sleep, and light sleep are reported in hours. Sleep quality was assessed via self-report on a 5-point scale with higher values indicating better sleep quality. The no-expectation-wake condition contains 18 participants (15 female), the no-expectation-sleep condition 14 participants (8 female), the expectation-wake condition 19 participants (14 female), and the expectation-sleep condition 23 participants (16 female). With the exception of 9 participants ($n_{\text{no expectation, wake}} = 2$, $n_{\text{no expectation, sleep}} = 3$, $n_{\text{expectation, wake}} = 3$, $n_{\text{expectation, sleep}} = 1$), all participants stated that they slept for at least 6 hr in the night before the first session of the experiment. All but one participant reported that they did not consume any alcoholic drinks between the first and second session.

[a] Assessed by a commercial fitness tracker (Xiaomi Mi Band 2). [b] How many times the sequence of learning phase, distractor task, and immediate cued recall task was repeated until participants responded at least 60% correctly on the immediate cued recall task. [c] $df = 34$, because sleep quality was not captured for one participant and the fitness tracker did not record any data for another participant.

Table S22

Means and Standard Errors for the Memory Performance Measures in Experiment 2 Including all Participants Previously Excluded due EMR Model Specific Exclusion Criteria

Dependent variable	No test expectation		Test expectation			
	Wake (*n* = 18)	Sleep (*n* = 14)	Wake (*n* = 19)	Sleep (*n* = 23)	*F*(1, 70)	*p*
Immediate cued recall	30.83 (0.51)	31.21 (0.56)	32.74 (0.56)	31.87 (0.43)	0.37	.547
Delayed cued recall	26.11 (0.74)	30.14 (0.60)	28.37 (0.78)	30.74 (0.54)	0.37	.543
Cued recall forgetting	4.72 (0.40)	1.07 (0.24)	4.37 (0.29)	1.13 (0.20)	0.12	.727
Free recall	14.06 (0.58)	16.00 (0.62)	14.37 (0.64)	15.52 (0.42)	0.12	.730

Note. The table shows the mean number of words (cued recall) and word pairs (free recall) correctly recalled in the respective memory test. Standard errors of the mean are presented in parentheses. Cued recall forgetting represents the difference between immediate and delayed cued recall performance. *F* = *F* value for the interaction of test expectation and wake versus sleep.

Additional Memory Performance Analysis for Experiment 2 Including all Participants Previously Excluded due EMR Model Specific Exclusion Criteria

Because we excluded four participants based on the application of the exclusion criteria recommended for the EMR model, we performed all ANOVAs for the memory performance analysis again, this time including the previously excluded participants (i.e., new sample size: $n_{\text{no expectation/wake}} = 18$, $n_{\text{no expectation/sleep}} = 14$, $n_{\text{expectation/wake}} = 19$, $n_{\text{expectation/sleep}} = 23$). As in the main analysis, none of our analyses showed a significant main effect of test expectation, delayed cued recall: $F(1, 70) = 1.10$, $MSE = 33.03$, $p = .297$, $\eta_p^2 = .02$; cued recall forgetting: $F(1, 70) = 0.06$, $MSE = 6.22$, $p = .803$, $\eta_p^2 < .001$; free recall: $F(1, 70) = 0.01$, $MSE = 23.44$, $p = .943$, $\eta_p^2 < .001$. Also in line with the main analysis, we found a statistically significant main effect of wake versus sleep with better memory performance after sleep compared to wakefulness in delayed cued recall, $F(1, 70) = 5.56$, $MSE = 33.03$, $p = .021$, $\eta_p^2 = .07$, and cued recall forgetting, $F(1, 70) = 34.19$, $MSE = 6.22$, $p < .001$, $\eta_p^2 = .33$. However, again no statistically significant main effect of wake versus sleep emerged in free recall: $F(1, 70) = 1.83$, $MSE = 23.44$, $p = .180$, $\eta_p^2 = .03$. Furthermore, consistent with the main analysis, no statistically significant interaction effects of test expectation and wake versus sleep emerged; delayed cued recall: $F(1, 70) = 0.37$, $MSE = 33.03$, $p = .543$, $\eta_p^2 = .01$; cued recall forgetting: $F(1, 70) = 0.12$, $MSE = 6.22$, $p = .727$, $\eta_p^2 = .002$; free recall: $F(1, 70) = 0.12$, $MSE = 23.44$, $p = .730$, $\eta_p^2 = .002$. Furthermore, the result patterns for immediate cued recall did not differ between the previous and current analyses; main effect of test expectation: $F(1, 70) = 1.54$, $MSE = 19.07$, $p = .219$, $\eta_p^2 = .02$; main effect of wake versus sleep: $F(1, 70) = 0.06$, $MSE = 19.07$, $p = .814$, $\eta_p^2 < .001$; interaction effect of test expectation and wake versus sleep: $F(1, 70) = 0.37$, $MSE = 19.07$, $p = .547$, $\eta_p^2 = .01$. Thus, using the larger sample size of $N = 74$, we observed similar results (for sample characteristics, sample means and standard errors of all experimental conditions, see Tables S21 and S22).

References

Gelman, A., & Rubin, D. B. (1992). Interference from iterative simulation using multiple sequences. *Statistical Science, 7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

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

Randler, C. (2013). German version of the reduced Morningness-Eveningness Questionnaire (rMEQ). *Biological Rhythm Research, 44*(5), 730–736. https://doi.org/10.1080/09291016.2012.739930

Does Sleep Benefit Source Memory?

Investigating 12-hr Retention Intervals with a Multinomial Modeling Approach

Sabrina Berres, Edgar Erdfelder, and Beatrice G. Kuhlmann

Department of Psychology, School of Social Sciences, University of Mannheim

Author Note

Sabrina Berres https://orcid.org/0000-0001-9868-8824

Edgar Erdfelder https://orcid.org/0000-0003-1032-3981

Beatrice G. Kuhlmann https://orcid.org/0000-0002-3235-5717

Hypotheses, study designs, sample sizes, and analysis plans were preregistered (Experiment 1: https://osf.io/gctzn; Experiment 2: https://osf.io/a6z4u). Additional materials (i.e., list of stimuli, codebook, data set) are provided online on the Open Science Framework (OSF; https://osf.io/8rmj2/?view_only=02e5eec5c3e54fd4aff3d55eedebffa7). We have no conflicts of interest to disclose. This research was funded by the University of Mannheim's Graduate School of Economic and Social Sciences (GESS) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grants KU3329/1-1 (Emmy Noether Research Group "Source Forgetting in Younger and Older Adults") and GRK 2277 ("Statistical Modeling in Psychology" [SMiP]). We thank

Sophia Laubel and Lena Hopf for their help in collecting data for Experiment 1 and Experiment 2, respectively.

Correspondence concerning this article should be addressed to Sabrina Berres, Edgar Erdfelder, or Beatrice G. Kuhlmann, Department of Psychology, School of Social Sciences, University of Mannheim, L13, 15-17, Room 425, 68161 Mannheim, Germany. Email: sberres@mail.uni-mannheim.de, erdfelder@uni-mannheim.de, or kuhlmann@uni-mannheim.de

Abstract

For retention intervals of up to 12 hr, the active systems consolidation hypothesis (Inostroza & Born, 2013; Klinzing et al., 2019) predicts that sleep compared to wakefulness strengthens the context binding of memories previously established during encoding. Sleep should thus improve source memory. We tested this prediction in two online source-monitoring experiments using intentionally learned pictures as items and incidentally learned screen positions and frame colors as source dimensions. In Experiment 1, we examined source memory by varying the spatial position of pictures on the computer screen. Multinomial modeling analyses revealed a significant sleep benefit in source memory. In Experiment 2, we manipulated both the spatial position and the frame color of pictures orthogonally to investigate source memory for two different source dimensions at the same time, also allowing exploration of bound memory for both source dimensions. The sleep benefit on spatial source memory replicated. In contrast, no source memory sleep benefit was observed for either frame color or bound memory of both source dimensions, probably as a consequence of a floor effect in incidental encoding of color associations. In sum, the results of both experiments show that sleep within a 12-hr retention interval improves source memory for spatial positions, supporting the prediction of the active systems consolidation hypothesis. However, additional research is required to clarify the impact of sleep on source memory for other context features and bound memories of multiple source dimensions.

Keywords: episodic memory, source memory, consolidation during sleep, binding, multinomial processing tree modeling

Does Sleep Benefit Source Memory?

Investigating 12-hr Retention Intervals with a Multinomial Modeling Approach

Episodic memory refers to memory for past events, experiences, or the source (context)[1] of information (e.g., location, time; Tulving, 2002). Empirical evidence from neuroimaging techniques such as functional magnetic resonance imaging (fMRI) points to a crucial role of the hippocampus in episodic memory (for reviews, see Eichenbaum et al., 2007; Mitchell & Johnson, 2009). Specifically, the hippocampus appears to bind the content of memories (i.e., item memory) to its unique context (i.e., source memory) during encoding.

Our present research addresses the role of sleep in these source binding processes. Almost a century of research in neuroscience and psychology has impressively shown that episodic memory is supported by sleep (for a recent meta-analysis, see Berres & Erdfelder, 2021). One mechanism assumed to underlie the sleep benefit in episodic memory is memory consolidation. As such, memory consolidation during sleep increases episodic memory storage by converting recently encoded and therefore labile memories into more stable long-term memory representations (Buzsáki, 1998; Diekelmann & Born, 2010; Dudai, 2004, 2012; Dudai et al., 2015; Klinzing et al. 2019; Rasch & Born, 2013). There are various theories that explain sleep benefits in episodic memory by memory consolidation, such as the sequential hypothesis[2] (Ambrosini & Giuditta, 2001; Giuditta, 2014; Giuditta et al., 1995) and the synaptic homeostasis hypothesis[3] (Cirelli & Tononi, 2015; Tononi & Cirelli, 2003, 2006, 2014, 2020). In the current work, we focus on memory consolidation as proposed by the active systems consolidation hypothesis (Born & Wilhelm, 2012; Diekelmann & Born, 2010; Feld & Born, 2017; Inostroza & Born, 2013; Klinzing et al., 2019; Rasch & Born, 2013).

---

[1] We use the terms source and context interchangeably. This is in line with the more general definition of source in the source-monitoring framework (Johnson et al., 1993) that comprises both context features (e.g., spatial position) and internal features (e.g., cognitive operations; Mitchell & Johnson, 2009).

[2] The sequential hypothesis states that memories are consolidated in two consecutive steps during slow-wave sleep (SWS) and rapid eye movement (REM) sleep (Ambrosini & Giuditta, 2001; Giuditta, 2014; Giuditta et al., 1995).

[3] The synaptic homeostasis hypothesis states that sleep renormalizes synaptic connections that were strengthened during wakefulness, thereby restoring cellular homeostasis (Cirelli & Tononi, 2015; Tononi & Cirelli, 2003, 2006, 2014, 2020).

This hypothesis is arguably “the currently most integrative account of sleep-dependent memory consolidation” (Klinzing et al., 2019, p. 1598), because it incorporates aspects of various consolidation theories—including the sequential and synaptic homeostasis hypothesis. Specifically, the active systems consolidation hypothesis states that during wakefulness, components of a memory representation (e.g., color, texture, odor of a fruit) are formed and distributed across various neocortical brain areas. In parallel, the hippocampus binds these components to a unique memory representation. During subsequent sleep, especially during slow-wave sleep (SWS), the hippocampal memory representation is replayed by reactivating specific neuronal firing patterns (Klinzing et al., 2019; Lewis & Durrant, 2011; O’Neill et al., 2010; Pfeiffer, 2020; Wilson & McNaughton, 1994). These local synaptic upscaling processes strengthen not only synaptic connections in the hippocampus and thus stabilize the hippocampal memory representation but also strengthen the separate components of the memory representation by triggering replay in the various neocortical brain areas. Simultaneously, global synaptic downscaling renormalizes the strength of synaptic connections across all cortical and subcortical areas by diminishing neuronal firing rates (Feld & Born, 2017; Klinzing et al., 2019). It is assumed that the combination of local synaptic upscaling and global synaptic downscaling in the hippocampus and neocortex results in a net strengthening of episodic context-bound hippocampal memory representations for relatively short retention intervals (e.g., 12 hr) and more gist-like decontextualized neocortical memory representations for longer retention intervals (e.g., 3 days; Klinzing et al., 2019). This assumption is supported by studies indicating a strengthening but no decontextualization of episodic memories within 10 to 12 hr after learning (e.g., Jurewicz et al., 2016; Lutz et al., 2017). In brief, according to the active systems consolidation hypothesis, sleep compared to wakefulness within a 12-hr retention interval should strengthen associations between the components of a memory representation that were previously established during encoding.

To investigate the sleep benefit in episodic memory, researchers have often used item-item associations such as word pairs as stimulus material (Diekelmann et al., 2009; Klinzing et al., 2019;

for a meta-analysis on single words and word pairs, see Berres & Erdfelder, 2021). By contrast, only a few studies investigated the sleep benefit using item-source associations (for a discussion of functional differences between item-item and item-source associations, see Mayes et al., 2007). In the following section, we review the rather mixed outcomes of item-source association studies conducted so far.

Overview of Research on Sleep Benefits in Source Memory

Using a split-night design, Rauchs et al. (2004) found better free recall performance for spatial positions (i.e., top vs. bottom) of words in a what-where-when task after sleep in the second half of the night (i.e., sleep predominantly characterized by rapid eye movement [REM] sleep) compared to wakefulness. In contrast, sleep-wake comparisons in the second half of the night for word-list associations (i.e., temporal source memory, “when” dimension) showed no significant differences. Correspondingly, the authors found no significant differences for sleep-wake comparisons in the first half of the night (i.e., sleep predominantly characterized by SWS) for spatial positions and lists. Furthermore, all sleep-wake comparisons for spatial positions and lists in the subsequent recognition test were insignificant. When comparing sleep deprivation in the first versus the second half of the night, the authors found better free recall performance for word positions after SWS deprivation than after REM sleep deprivation (Rauchs et al., 2004). These results suggest that REM sleep contributes to the sleep benefit in item-position associations, thereby conflicting with the active systems consolidation hypothesis which considers SWS to be more important for memory consolidation. However, in line with the consolidation hypothesis, other split-night studies showed worse memory of the frame color and spatial position for neutral pictures after SWS deprivation than after REM sleep deprivation, pointing to a pivotal role of SWS for memory performance (see Groch et al., 2015; Sopp et al., 2017).

The results were also mixed for studies comparing naps versus wakefulness during the day or early evening: Wang and Fu (2009) as well as Köster et al. (2017) found no significant differences between sleep and wakefulness for picture-background color associations, contradicting the active systems consolidation hypothesis. By contrast, van der Helm et al. (2011) found a significant sleep benefit in source memory for word-context associations in line with the active systems consolidation hypothesis. Further support is provided by Lewis et al. (2011, Experiment 2), who observed significantly less forgetting after sleep than wakefulness in source memory for object-background photo associations.

Classical sleep study designs that compared nighttime sleep and daytime wakefulness using retention intervals up to 12 hr resulted in somewhat stronger evidence for sleep-induced context memory improvements as predicted by the active systems consolidation hypothesis. Lewis et al. (2011) made use of such a design in their first experiment and found results very similar to their nap study in Experiment 2. Also using a retention interval of 12 hr filled with either sleep or wakefulness, Mawdsley et al. (2014) observed a significant sleep benefit in source memory for word-position associations. Wang et al. (2017) investigated the sleep benefit for word pair-temporal context associations in children. Specifically, children learned two lists of word pairs with 1 hr between learning of the first and second list (temporal context). After a retention interval of 11 hr, memory for word pairs was tested with a cued recall task. In addition, children were asked to indicate the list of the respective word pair. The authors found that interpolated sleep compared to wakefulness improved memory for word pairs and the temporal context but not for word pair-temporal context associations, which contradicts the active systems consolidation hypothesis (Wang et al., 2017).

Overall, the empirical evidence concerning sleep benefits in source memory is thus quite mixed. The reviewed studies differ in several aspects that may explain the mixed results observed with respect to the sleep benefit in item-context associations. For example, researchers have not only used a wide variety of sleep study designs (i.e., split-night designs, daytime naps, nighttime naps, natural sleep and wakefulness), but also different stimulus materials for items (i.e., single words, word pairs, pictures) and sources (i.e., spatial positions, frame colors, background colors, background photos, posters, lists), next to different encoding instructions (i.e., intentional learning of item-context associations, incidental learning of item-context associations, intentional learning of items but incidental learning of contexts).

Further important, the source memory measures used likely contribute to the inconsistent results pattern. According to the source-monitoring framework, multiple cognitive processes such as memory, decision making, guessing, and response biases are involved in making judgments about the origin of a memory (Johnson et al., 1993). These cognitive processes are confounded in frequently used standard measures of source memory (cf. Batchelder & Riefer, 1990). Source memory is often measured by simply counting the number of correct source attributions (e.g., Groch et al., 2015; Lewis et al., 2011; Mawdsley et al., 2014; Wang et al., 2017) or by using the source identification measure (SIM; e.g., van der Helm et al., 2011), defined as the proportion of correct source attributions for all target items, irrespective of whether they were identified as “old” or “new”. Another frequently used measure for source memory is the average conditional source identification measure (ACSIM; Rauchs et al., 2004; Sopp et al., 2017; Wang & Fu, 2009), defined as the proportion of correct source attributions for all target items correctly identified as “old”, averaged across the two sources (e.g., left, right). Although item and source memory are somewhat less confounded in ACSIM than in SIM, all of the listed source memory measures confound item memory, source memory, and guessing to some degree (Bröder & Meiser, 2007; Murnane & Bayen, 1996). We therefore argue that more rigorous and less contaminated measures of source memory are required to test whether sleep compared to wakefulness strengthens the context binding of episodic memories for retention intervals up to 12 hr, as predicted by the active systems consolidation hypothesis (Inostroza & Born, 2013; Klinzing et al., 2019).

The Current Experiments

We tested the sleep-strengthens-source-memory hypothesis in two experiments using multinomial processing-tree models to disentangle cognitive processes that are likely confounded in commonly used measures of source memory. In Experiment 1, we manipulated the spatial position of pictures on a computer screen in a standard source-monitoring task (e.g., Bayen et al., 1996;

Murnane & Bayen, 1996) to investigate source memory for item-context associations after a 12-hr retention interval filled with either a period of nighttime sleep or daytime wakefulness. We conducted a second experiment with the main purpose to conceptually replicate the results for spatial position memory of Experiment 1. In Experiment 2, we additionally manipulated the frame color orthogonally to the spatial position of pictures. This allowed us to explore two additional research questions: First, can the results for spatial position memory be generalized to other source dimensions (i.e., frame color memory)? Second, does sleep compared to wakefulness benefit memory for context-context associations (i.e., bound source memory for spatial position and frame color)?

Methodological Considerations

In both experiments, we explicitly instructed participants to study the pictures for a later recognition test (i.e., intentional learning of items), whereas no such instruction was provided for their sources (i.e., sources were learned incidentally). To counteract possible floor effects in source memory, participants performed an orienting task during the learning phase that requires attending to the relevant source information but involves no rehearsal (i.e., indicating spatial positions using response keys; cf. Boywitt & Meiser, 2012). By preventing participants from using explicit rehearsal strategies for item-context and context-context associations, this approach creates a more realistic setting for examining everyday source monitoring. Note that most previous studies on the sleep benefit concerning context-binding employed intentional learning of item-context associations (for incidental learning, for example, see Mawdsley et al., 2014; Wang et al., 2017).

To allow comparisons with previous studies, we report hit rates and false-alarm rates in addition to the sensitivity index d' and response bias c for item memory. Whereas sensitivity and response bias are confounded in hit rates (i.e., proportion of target items correctly identified as “old”) and false-alarm rates (i.e., proportion of distractor items falsely identified as “old”), sensitivity and response bias are separated in d' and c as derived from the signal detection theory (SDT; Stanislaw & Todorov, 1999; e.g., van der Helm et al., 2011). Specifically, larger positive values of d' indicate better discrimination between target and distractor items. Response bias *c* denotes the general tendency to respond “old” or “new”, with larger negative values indicating a stronger “old”-response bias, values close to zero no response bias, and larger positive values a stronger “new”-response bias (Stanislaw & Todorov, 1999).

For source memory, we report the average conditional source identification measure (ACSIM), defined as the proportion of correct source attributions for all target items correctly identified as “old”, averaged across the two sources (e.g., left, right) of a source dimension (e.g., spatial position; Murnane & Bayen, 1996). Because ACSIM is not defined when all target items correctly identified as “old” are assigned to the same source (e.g., right) of a source dimension (e.g., spatial position), we report the conditional source identification measure (CSIM) in these cases. This measure is defined as the averaged proportion of correct source attributions for all target items correctly identified as “old”. For ACSIM and CSIM, larger positive values indicate better source memory. Note, however, that both measures confound source memory with item memory in some circumstances, for example, when targets are identified as “old” based on guessing (Bayen et al., 1996; for a detailed discussion, see Murnane & Bayen, 1996).

In contrast to ACSIM and CSIM, multinomial processing tree (MPT) models allow to disentangle source memory from item memory and guessing (for reviews on this model class and a MPT tutorial, see Batchelder & Riefer, 1999; Erdfelder et al., 2009; Schmidt et al., 2023). MPT models have therefore gained considerable popularity in source memory research (e.g., Arnold et al., 2019; Bell et al., 2017; Boywitt & Meiser, 2012; Kuhlmann et al., 2016). There are several options for fitting MPT models to empirical data (e.g., Heck et al., 2018; Moshagen, 2010), with complete and partial pooling being the two most often used methods. Specifically, in the complete pooling approach, observed category frequencies are aggregated across participants, and the maximum likelihood (ML) method is used to obtain MPT-parameter estimates. In contrast to complete pooling, the partial pooling approach explicitly accounts for potential individual differences between participants by combining information on the individual and group level. For individual and group-level parameter estimation, partial pooling relies on a Bayesian approach employing Markov-chain Monte Carlo (MCMC) methods (Heck et al., 2018). Here we used both methods to check whether our results are robust against the different distributional assumptions involved in complete and partial pooling. We used the software multiTree (Moshagen, 2010) and the latent-trait approach (Klauer, 2010) as implemented in the R package TreeBUGS (Heck et al., 2018) for complete and partial pooling, respectively.

Transparency and Openness

Hypotheses, study design, sample size, and analysis plan were preregistered for Experiment 1 (https://osf.io/gctzn) and Experiment 2 (https://osf.io/a6z4u). For both experiments, the data and stimulus materials are available at the Open Science Framework (OSF; https://osf.io/8rmj2/?view_only=02e5eec5c3e54fd4aff3d55eedebffa7). In the respective Method sections, we provide detailed information about the MPT models used, sample size determination, and data exclusions.

Experiment 1

To reiterate, according to the active systems consolidation hypothesis, the hippocampus binds the content (i.e., item memory) and its unique context (i.e., source memory) to a unique memory representation during encoding. This memory representation is replayed during subsequent sleep which should result in better item and source memory compared to wakefulness (Feld & Born, 2017; Klinzing et al., 2019). For a 12-hr retention interval, the active systems consolidation hypothesis thus predicts that both item memory and source memory should benefit from sleep. To test these two hypotheses, we used the two-high-threshold MPT model of source monitoring (2HTSM) shown in Figure 1. The 2HTSM model performed best in a comparative validation study of source monitoring models (Bayen et al., 1996). As such, this model is based on a standard source-monitoring task in which participants study items from two sources and are subsequently asked whether the item was previously presented, and if so, in which source (e.g., Bayen et al., 1996; Murnane & Bayen, 1996). The 2HTSM provides separate parameters for item memory, source memory, and guessing.

Specifically, participants correctly recognize a target item presented in source A or B as "old" or a distractor item as "new" with probability *D*. Conditionally on correct item recognition, participants correctly identify the source with probability *d*. However, if item memory (1 - *D*) or source memory (1 - *d*) fails, participants are assumed to guess. In case of successful item memory but failing source memory, participants correctly guess the source of a target item with probability *a*. If item memory fails, participants guess "old" with probability *b*. Finally, if both item and source memory fail, participants correctly guess the source with probability *g* (Bayen et al., 1996).

In the most general version of the 2HSTM, item memory, source memory, and source guessing may vary between item types and sources as illustrated in Figure 1. To arrive at an identifiable and most parsimonious 2HTSM submodel that still fits the data, we first tested invariance of item memory with respect to item types and sources, followed by invariance tests of source memory, and finally guessing. By using this principled strategy, we aimed at identifying a submodel with a minimum of precisely estimable parameters (see Bayen et al., 1996).

According to the active systems consolidation hypothesis, the corresponding item memory and source memory parameters should both be larger when participants sleep during the 12-hr retention interval than when they stay awake.

Method

In this experiment, we compared participants randomly assigned to a wake versus sleep condition. Whereas participants in the wake condition learned the material in the morning and were tested in the evening after a 12-hr retention interval of daytime wakefulness, this was reversed for participants in the sleep condition who were tested after a period of nighttime sleep. Crucially note that previous research showed comparable performance in learning as well as testing parameters by using the same sleep study design, showing that circadian effects are not a serious confound in this design (e.g., Abel & Bäuml, 2012, 2013a, 2013b, 2014; Bäuml et al., 2014; Erdfelder et al., 2022; Fenn & Hambrick, 2013).

Participants

We determined the necessary sample size a priori by conducting two power analyses: First, despite our directional predictions, we conservatively performed an a priori power analysis for a two-tailed *t* test with two independent groups using G*Power 3.1 (Faul et al., 2007). Given a medium effect size (Cohen's **d** = 0.50), a conventional α-level of .05, and a target-power of 1 - β = .80, the analysis resulted in a total sample size of 128 participants. Second, we determined the necessary sample size for the model-based analysis using multiTree (Moshagen, 2010). Assuming a sleep-wake difference of .10 in the crucial parameter (*D* or *d*, depending on the hypothesis), an analysis based on 130 participants, 60 target items, and 30 distractor items resulted in a power larger than .99 for the item memory parameter *D* and a power of .96 for the source memory parameter *d* (for more detailed information, see the preregistration on the OSF, https://osf.io/gctzn). Thus, we strove for a sample of 130 participants. Data collection took place from fall 2020 to spring 2021. Note that we extended the data collection phase until we reached the desired number of participants because data collection was slow and only a fraction of the targeted sample size was collected within the preregistered three months.

In total, 174 participants recruited via mailing lists of the University of Mannheim, social media, personal contacts, and the online research platform Prolific (https://www.prolific.co; Palan & Schitter, 2018; Peer et al., 2017) took part in the online experiment. After successful completion of the experiment, 103 participants recruited via Prolific (59.20%) were paid a flat fee of £ 4.50, whereas 71 participants recruited through other channels (40.80%) either received corresponding course credits or were eligible to win vouchers. Due to random assignment to the wake versus sleep condition, the number of participants who were paid (n_{wake} = 50, n_{sleep} = 53), received corresponding course credits or were eligible to win vouchers (n_{wake} = 40, n_{sleep} = 31) were approximately balanced across the experimental conditions.[4] Note that the experiment was successfully completed only if

---

[4] To check whether our results are confounded by the type of compensation, we conducted additional exploratory analyses based on the analyzed data of Experiment 1 and Experiment 2. Results are provided in the Supplemental Materials (see Table S2 and S5). There was no evidence that item memory (i.e., hit rates, *d'*) and ACSIM-based source memory performance in the wake and sleep conditions were affected by the type of compensation.

the following two conditions were met: First, all parts of the experiment had to be completed within the set time frames (i.e., registration, learning, and testing session). Second, more than 50% of the responses in the orienting task had to be correct.

Following the preregistered exclusion criteria, 23 participants were excluded from the analysis, because they indicated that they were distracted or disturbed during the experiment. Another four participants had to be excluded because the retention interval was not within 11–13 hr. Furthermore, seven participants of the wake condition were excluded because they napped during the retention interval, and two participants were excluded because they reported having neurological disorders. We also excluded two participants because of substantial alcohol consumption (i.e., females were excluded if they consumed more than 20 g alcohol, males were excluded if they consumed more than 40 g alcohol), and one participant with a larger false-alarm rate than hit rate. Three additional participants were excluded for unforeseen reasons not included in the preregistration: One participant reported using memory aids (e.g., notes, screenshots), one participant reported technical problems, and another participant assigned to the wake condition delayed the start of the experiment so that it started in the evening instead of the morning. In sum, we excluded 42 participants, leaving 132 participants for analysis, all of them fluent in German. The 132 participants were between 18 and 35 years of age (*M* = 26.77 years, *SD* = 4.48), 84 (63.64%) were female (see Table S1 in the Supplemental Materials for more detailed sample characteristics).

Materials

We selected 160 colored object photos from the bank of standardized stimuli (BOSS; Brodeur et al., 2010) of which 60 randomly chosen target pictures were displayed on either the left or right side of the screen (i.e., 30 pictures each were displayed at the 10% and 90% position on the x-axis). Thus, spatial positions of pictures (left vs. right) served as the two sources of interest. Another 30 pictures were randomly selected as distractors, and four additional pictures were randomly selected as buffer items which were included at the start of the learning phase to prevent primacy effects. Note that we decided against including a recency buffer because of the 12-hr retention interval. A list of the 160 pictures and detailed information about the selection criteria are available at the OSF (https://osf.io/8rmj2/?view_only=02e5eec5c3e54fd4aff3d55eedebffa7).

Procedure

The online experiment was conducted with SoSci Survey (Leiner, 2020), using lab.js (Henninger et al., 2022) for stimulus presentation during the study phase, and consisted of three parts: registration, learning, and testing session. In the registration session, participants gave informed consent before being randomly assigned to either the sleep or wake condition. They were asked to pick a date and time for the first session in line with their randomly predetermined condition (i.e., wake condition: 7 a.m. to 10 a.m.; sleep condition: 7 p.m. to 10 p.m.) and were informed that the second session starts 12 hr later. Participants received the access link via email or Prolific notification 15 min before the start of the learning session. During the study phase, 64 randomly selected pictures (i.e., 4 buffer and 60 target items) were sequentially presented on the left or right side of the screen for 4 s each with an interstimulus interval of 1 s (i.e., blank white screen for 500 ms followed by a fixation cross for 500 ms). While a picture was presented on the screen, participants performed the orienting task, which entailed pressing the correct button for the spatial position. The two buttons labeled “left” and “right” were arranged next to each other and were displayed below the picture. Only participants who answered with the correct spatial position for more than 50% of the 64 pictures completed the learning session and were invited to the testing session 12 hr later. Again, participants received the access link via email or Prolific notification 15 min before the session started. For the testing session, the 60 target items were intermixed with 30 distractor items and presented in the middle of the screen with two buttons labeled “old” and “new” below. Note that we varied the spatial position of the labels “old” and “new” randomly between participants but kept it constant within participants. By pressing one of the two buttons, participants indicated whether the picture was presented during the study phase (“old”) or not (“new”). If participants answered “old”, they were asked whether the picture was presented left or right and to respond with the corresponding button. This task was followed by control and demographic questions (for details, see the preregistration on the OSF, https://osf.io/gctzn), before participants were finally thanked and debriefed.

Results

We set a significance level of $\alpha = .05$ for all analyses. For hit and false alarm rates as well as *d′* and *c* measures of item recognition we report means, standard errors, and *t*-test results in Table 1.[5] Regarding item memory, all two-tailed *t* tests for two independent groups showed insignificant differences between the sleep and wake condition, $t(130) \leq 1.62$, $p \geq .107$ (see Table 1). In contrast, source memory as measured by ACSIM significantly benefitted from sleep, $t(130) = 3.46$, $p = .001$, estimated Cohen's **d** = 0.59 (sleep condition: *M* = 0.77, *SE* = 0.01; wake condition: *M* = 0.69, *SE* = 0.01). Taken together, using commonly applied measures of item and source memory, we found statistically significant evidence for a sleep benefit in source memory but not in item memory.

The most parsimonious model we originally aimed at—Submodel 4 of the 2HTSM with parameter *D* for item memory, parameter *d* for source memory, and parameters *b* and *g* for guessing (Bayen et al., 1996; see the preregistration on the OSF, https://osf.io/gctzn)—produced considerable misfit for the aggregated data, $G^2(4) = 10.21$, $p = .037$. While invariance of item and source memory parameters across item types and sources turned out to be unproblematic, assuming invariance of source guessing parameters *a* and *g* in addition resulted in the observed misfit. Hence, applying Submodel 5a of the 2HTSM (Bayen et al., 1996)—with a single parameter *D* for item memory, a single parameter *d* for source memory, and three parameters *a*, *b*, and *g* for guessing—resulted in a good fit, $G^2(2) = 1.78$, $p = .411$. The ML parameter estimates, standard errors, and 95% confidence intervals of Submodel 5a for the wake and sleep condition are summarized in Table 2. We found a statistically significant difference between sleep versus wake conditions in the item memory parameter *D*, $\Delta G^2(1) = 13.66$, $p < .001$. The item memory parameter estimate for the sleep condition

---

[5] In both experiments, we calculated the sensitivity index *d′* by $Z_{\text{hit rate}} - Z_{\text{false-alarm rate}}$ and response bias *c* by $-0.5 * (Z_{\text{hit rate}} + Z_{\text{false-alarm rate}})$, with *Z* denoting the probit transformation (see, for example, Stanislaw & Todorov, 1999). Note that we used the log-linear rule (i.e., adding 0.50 to hits, false alarms, misses, and correct rejections; Hautus, 1995) to obtain estimates of *d′* and *c* even in case of extreme hit rates and false alarm rates of 0 or 1.

was almost 5% larger than for the wake condition. Similarly, the source memory parameter *d* also differed significantly between conditions, $\Delta G^2(1) = 31.30$, $p < .001$, with about 15% higher source memory estimates after sleep than after wakefulness. Concerning the guessing parameters, we found significantly more "old"-guessing in the wake than the sleep condition (parameter *b*), $\Delta G^2(1) = 6.09$, $p = .014$; and a significantly stronger "left" guessing bias for unrecognized items after sleep than after wakefulness (parameter *g*), $\Delta G^2(1) = 10.13$, $p = .001$. By contrast, there was no statistically significant difference between the sleep and wake condition in source guessing for recognized items (parameter *a*), $\Delta G^2(1) = 1.36$, $p = .243$.

To check the robustness of our results, we reanalyzed the same data in the framework of Klauer's (2010) hierarchical latent-trait model as implemented in TreeBUGS (Heck et al., 2018). As can be seen in the Appendix (see Table A1), the estimated group-level means resembled those reported in Table 2. We thus conclude that the basic result pattern does not depend on whether complete or partial pooling approaches are used for data analysis.

Discussion

Both the ACSIM-based and the model-based results suggest that sleep compared to wakefulness benefits source memory. This is in line with a core prediction of the active systems consolidation hypothesis that sleep benefits source memory for retention intervals of up to 12 hr.

For item memory, the descriptive result patterns of *d'* and the aggregated as well as hierarchical model-based analyses suggest that sleep compared to wakefulness might benefit item recognition. Whereas item memory was descriptively higher after sleep versus wakefulness in all three analyses, the sleep benefit was significant only for complete pooling. This deviance is likely due to different analysis-levels (i.e., complete pooling, partial pooling, no pooling) that account for potential individual differences to a varying extend. Our results are in line with previous research that uses recognition tasks to assess item memory, yielding mixed evidence for the active systems consolidation hypothesis: Some studies found a significant sleep benefit in item memory (e.g., Köster et al., 2017; Mawdsley et al., 2014; Wang et al., 2017), whereas others did not (e.g., van der

Helm et al., 2011; Wang & Fu, 2009). In fact, a recent meta-analysis showed that the sleep benefit for word materials is largest in free recall, followed by cued recall, and lastly recognition tasks (Berres & Erdfelder, 2021). This suggests that item recognition apparently benefits from sleep only slightly, thereby making it difficult to detect these small positive sleep effects in item recognition tasks (e.g., Rauchs et al., 2004; Wang & Fu, 2009).

In sum, Experiment 1 indicates that sleep improves source memory within a 12-hr retention interval as predicted by the active systems consolidation hypothesis. However, to establish the validity of this conclusion more rigorously, our results require an experimental follow-up evaluation. We therefore conducted a second experiment with the aim to conceptually replicate the results for spatial position memory. By manipulating frame color orthogonally to the spatial position of pictures in Experiment 2, we were able to additionally explore whether the results for spatial position memory extent to a second source dimension (i.e., frame color). Furthermore, we explored whether sleep within a 12-hr retention interval also strengthens bound memory for spatial position and frame color.

Experiment 2

As in Experiment 1, we predict that both item memory and source memory should benefit from sleep compared to wakefulness in a 12-hr retention interval. Because hippocampal memory representations include not only item-context but also context-context associations, we also explored whether sleep improves bound memory for two source dimensions. We tested these predictions using a reparameterized variant of the MPT model of multidimensional source monitoring (Meiser, 2014), shown in Figure 2. Like the 2HTSM, this model is based on a source-monitoring task that is, however, extended to two source dimensions (e.g., a position dimension with sources “left” and “right”, and a color dimension with sources “blue” and “yellow”; Meiser, 2014).

The multinomial model of multidimensional source monitoring provides separate parameter estimates for item memory, bound source memory (i.e., spatial position plus frame color), unbound source memory (e.g., spatial position only), and guessing. Specifically, participants correctly recognize a target item presented by source *i* of the first source dimension (e.g., “left” or “right” on source dimension “spatial position”) and source *j* of the second source dimension (e.g., “blue” or “yellow” on source dimension “frame color”) as “old” with probability D_{ij} or detect a distractor item as “new” with probability D_{new}. Conditionally on correct item recognition, participants correctly identify the source combination (e.g., left and blue, left and yellow, right and blue, right and yellow) of recognized items with bound source probability d_{ij}. In contrast, if bound source memory fails for recognized items (i.e., the source combination is not correctly identified with probability $1 - d_{\text{ij}}$), participants can still correctly identify the sources *i* (e.g., “left” or “right” on source dimension “spatial position”) and *j* (e.g., “blue” or “yellow” on source dimension “frame color”) of either or both source dimensions independently with probabilities $e_{\text{ij}}^{\text{Position}}$ and $e_{\text{ij}}^{\text{Color}}$, respectively. However, if item memory ($1 - D_{\text{ij}}$), bound source memory ($1 - d_{\text{ij}}$), and unbound source memory ($1 - e_{\text{ij}}^{\text{Position}}$, $1 - e_{\text{ij}}^{\text{Color}}$) fail, participants are assumed to guess. In case of successful item memory but bound-source-memory and unbound-source-memory failure for either or both source dimensions, participants guess source A of source dimension *i* (e.g., “left” on source dimension “spatial position”) for a target item with probability a^{Position}. They also guess source X of source dimension *j* (e.g., “blue” on source dimension “frame color”) for a target item assigned to source A (e.g., left) or B (e.g., right) of source dimension *i* (e.g., spatial position) with probability $a_{|\text{left}}^{\text{Color}}$ or $a_{|\text{right}}^{\text{Color}}$, respectively. If item memory fails, participants guess “old” with probability *b*. For unrecognized target or distractor items identified as “old”, participants guess source A of source dimension *i* (e.g., “left” on source dimension “spatial position”) with probability g^{Position}. In addition, they guess source X of source dimension *j* (e.g., “blue” on source dimension “frame color”) for unrecognized target or distractor items assigned to source A (e.g., left) or B (e.g., right) of source dimension *i* (e.g., spatial position) with probability $g_{|\text{left}}^{\text{Color}}$ or $g_{|\text{right}}^{\text{Color}}$, respectively (Meiser, 2014).

In its most general version, the multidimensional source memory model allows for parameters that may differ between item types and sources as illustrated in Figure 2. To simplify this model and ensure identifiability of parameters, we employed basically the same principled strategy as previously used for the 2HTSM in Experiment 1. Specifically, we successively imposed the following set of constraints on the parameters (cf. Meiser, 2014; Meiser & Bröder, 2002): First, the item memory parameters D_{ij} were equated across the source dimensions “spatial position” and “frame color”, and D_{New} was constrained to be equal to the resulting item memory parameter D. Second, the bound source memory parameters d_{ij} were also equated across the source dimensions “spatial position” and “frame color” (parameter d). Next, the unbound source memory parameters for spatial position $e_{\text{ij}}^{\text{Position}}$ and frame color $e_{\text{ij}}^{\text{Color}}$ were equated across the source dimension “frame color” (parameter e^{Position}) and “spatial position” (parameter e^{Color}), respectively (Meiser, 2014; Meiser & Bröder, 2002). Finally, additional equality constraints were imposed on the source guessing parameters (i.e., $a^{\text{Position}} = g^{\text{Position}}$, $a_{|\text{left}}^{\text{Color}} = g_{|\text{left}}^{\text{Color}}$, $a_{|\text{right}}^{\text{Color}} = g_{|\text{right}}^{\text{Color}}$).

Drawing on the active systems consolidation hypothesis we predict for a 12-hr retention interval that the corresponding item memory parameters, bound source memory parameters, and unbound source memory parameters e^{Position} and e^{Color} should be larger after sleep than wakefulness.

Method

A 2 × 2 mixed factorial design with source dimension (spatial position vs. frame color) as within-subject factor and wake versus sleep as between-subjects factor was used in this experiment. Like in Experiment 1, participants were randomly assigned to a wake or sleep condition and learned the material either in the morning or in the evening before they were tested following a 12-hr retention interval.

Participants

To determine the necessary sample size for the model-based analysis a priori, we used multiTree (Moshagen, 2010). For an α-level of .05 and an assumed difference of .10 in the parameter of interest between the sleep and wake condition, the analysis for 130 participants, 120 target items, and 60 distractor items resulted in a power larger than .99 for item memory D, a power of .78 for bound source memory d, and power values of .67 and .61 for unbound source memories e^{Position} and

e^{Color}, respectively (for more detailed information, see the preregistration on the OSF, https://osf.io/a6z4u). As already detailed for Experiment 1, we aimed at a sample size of 130 participants in Experiment 2 and extended the pre-registered data collection period for the same reason as in Experiment 1. Specifically, we collected data from fall 2020 to spring 2021, using the same channels for participant recruitment as in Experiment 1. We made sure that participants of Experiment 1 did not additionally participate in Experiment 2 and vice versa.

In total, 175 participants took part in the online experiment and were rewarded for successful completion with a flat fee of £ 6.00, as Experiment 2 took longer to complete than Experiment 1. Following the preregistered exclusion criteria, 14 participants were excluded because they indicated that they were distracted or disturbed during the experiment. One participant admitted not to have taken the testing session seriously and was thus excluded. Two further participants were excluded because the retention interval was not within 11–13 hr. In addition, we excluded eight participants of the wake condition who napped during the retention interval. Another two participants reported having neurological disorders and were thus excluded. Furthermore, we excluded two participants because of substantial alcohol consumption[6], and three participants with larger false-alarm rates than hit rates.[7] We also excluded nine additional participants for unforeseen reasons not included in the preregistration: Four participants reported technical problems, and one participant assigned to the wake condition delayed the start of the experiment so that it started at noon instead of the morning. Four additional participants were excluded because they indicated having detailed knowledge about the study design or the aim of the experiment. In sum, we excluded 41 participants, leaving 134 participants for analysis, all of them fluent in German. These 134 participants were between 18 and 35 years of age (*M* = 25.58 years, *SD* = 4.51), 84 (62.69%) were female (see Table S4 in the Supplemental Materials for more detailed sample characteristics).

---

[6] When excluding all six participants who consumed alcohol ($n_{\text{wake}} = 4$, $n_{\text{sleep}} = 2$), the result patterns for item and source memory performance did not change (see Table S6 in the Supplemental Materials).

[7] One participant met the preregistered exclusion criterion “color blindness”. However, the participant reported being red-green color blind. This does not affect discrimination of blue and yellow. Therefore, the participant was not excluded from the analysis.

Materials

The stimulus material consisted of 200 grey-scaled object drawings selected from the multilingual picture databank (MultiPic; Duñabeitia et al., 2018). Of the 200 drawings, 120 target pictures were randomly chosen for each participant and displayed on either the left or the right side of the screen (i.e., 30 pictures each were displayed at the 10% and 90% position on the x-axis) with either a blue or a yellow colored frame (i.e., 30 pictures each were displayed in a blue colored 20-px frame with red-green-blue [RGB] values of 0, 40, 255, and in a complementary gold colored 20-px frame with RGB values of 255, 215, 0). Hence, spatial position (left vs. right) and frame color (blue vs. yellow) served as the two source dimensions of interest. Note that each source combination appeared equally often (i.e., 30 times). Another 60 pictures were randomly selected as distractors. Finally, four additional pictures were randomly selected as buffer items and presented in the beginning of the learning phase to prevent primacy effects. Hence, responses to these items were not included in our data analyses. As in Experiment 1, no recency buffer was included because of the 12-hr retention interval. A list of the 200 pictures and detailed information about the selection criteria are available at the OSF (https://osf.io/8rmj2/?view_only=02e5eec5c3e54fd4aff3d55eedebffa7).

Procedure

The procedure followed that of Experiment 1 and used the same online study builders (i.e., SoSci Survey, lab.js), but with the following changes to the source-monitoring task: During the study phase, 124 randomly selected pictures (i.e., 4 buffer and 120 target items) were sequentially presented on the left or right side of the screen in a blue or yellow colored frame for 5 s each with an interstimulus interval of 1 s (i.e., blank white screen for 500 ms followed by a fixation cross for 500 ms). The orienting task entailed pressing the correct button for spatial position and frame color of a picture shown on the screen. The two buttons for spatial position labeled “left” and “right” were arranged next to each other and were displayed in gray on the left side below the picture. The two buttons for frame color labeled “blue” and “yellow” were also arranged next to each other but displayed in the respective color on the right side below the picture. Only participants who answered with the correct spatial position for more than 50% of the 124 pictures completed the learning session and were invited to the testing session 12 hr later. For the testing session, the 120 target items were intermixed with 60 distractor items and presented frameless in the middle of the screen with two buttons labeled “old” and “new” below. Note that the spatial position of the labels “old” and “new” was varied between participants as in Experiment 1. If participants answered “old”, they were asked about the spatial position and frame color of the picture. To respond, participants pressed one of the two respectively labeled left buttons for “left plus blue” or “left plus yellow” or one of the two respectively labeled right buttons for “right plus blue” or “right plus yellow”. Whereas the labels for spatial position were always presented on the respective sides of the screen, the position for the frame-color labels was varied between participants (i.e., for half of the participants the labels “left plus blue” and “right plus blue” were displayed above “left plus yellow” and “right plus yellow”, while this order was reversed for the other half of the participants; for a detailed description of the procedure, see the preregistration on the OSF, https://osf.io/a6z4u).

Results

As in Experiment 1, we first analyzed effects of sleep versus wakefulness on commonly used measures of item and source memory. Again, we set a significance level of $\alpha = .05$ for all analyses. Means, standard errors, and *t*-test results for hit and false alarm rates as well as *d′* and *c* of item recognition are reported in Table 3. Concerning item memory, all two-tailed *t* tests between the sleep and wake groups were insignificant, $t(132) \leq 1.29$, $p \geq .199$ (see Table 3). To analyze ACSIM-based source memory, we performed a mixed ANOVA using source dimension (spatial position vs. frame color) as within-subject factor and wake versus sleep as between-subjects factor. There was a statistically significant main effect of the wake versus sleep condition, with better source memory after sleep ($M = 0.61$, $SE = 0.01$) than after wakefulness ($M = 0.57$, $SE = 0.01$), $F(1, 132) = 6.71$, $p = .011$, $\eta_p^2 = .05$. We also found a statistically significant main effect of source dimension, with better source memory for spatial position ($M = 0.64$, $SE = 0.01$) than for frame color ($M = 0.55$, $SE = 0.01$), $F(1, 132) = 42.59$, $p < .001$, $\eta_p^2 = .24$. However, there was no statistically significant interaction effect of the wake versus sleep condition and source memory, $F(1, 132) = 2.97$, $p = .087$. In sum, using commonly used item and source memory measures, we found a statistically significant sleep benefit in source memory but not in item memory.

In a second step, we tested our hypotheses using the MPT model of multidimensional source monitoring (Meiser, 2014) as described above. First, we fitted the most parsimonious model version, including equality constraints on the source guessing parameters (i.e., $a^{\text{Position}} = g^{\text{Position}}$, $a_{|\text{left}}^{\text{Color}} = g_{|\text{left}}^{\text{Color}}$, $a_{|\text{right}}^{\text{Color}} = g_{|\text{right}}^{\text{Color}}$). Applying this model to the aggregated data, however, resulted in misfit, $G^2(24) = 41.42$, $p = .015$. As in case of the 2HTSM used in Experiment 1, we therefore fitted the model without the additional constraints on the source guessing parameters to the aggregated data, which resulted in a good fit, $G^2(18) = 23.82$, $p = .161$. The ML parameter estimates, standard errors, and 95% confidence intervals of this model version for the four experimental conditions are displayed in Table 4. We found a statistically significant difference between the sleep and wake condition in item memory D, $\Delta G^2(1) = 31.54$, $p < .001$, with about 6% larger item recognition estimates following sleep. This matches the result for item memory observed in the aggregate analyses of Experiment 1. By contrast, there was no statistically significant sleep benefit in bound source memory d, $\Delta G^2(1) = 0.003$, $p = .955$. The likely reason for this unexpected result is that in both the sleep and the wake condition the estimate of correctly identifying the source combination of recognized target items was very low ($d = .04$), reflecting a floor effect in episodic context-context binding. Because bound source memory is very low, the unbound source memory parameters e^{Position} and e^{Color} resemble the source memory parameter d for item-context binding of the 2HTSM in Experiment 1.[8] Replicating Experiment 1, the unbound source memory parameter for spatial position e^{Position} differed significantly between the sleep and wake condition, $\Delta G^2(1) = 7.93$, $p = .005$. In line with the active systems consolidation hypothesis, the item-context binding probability for the source

---

[8] For this reason, we refrain from reporting the preregistered 2HTSM-analyses for separate source dimensions. Not surprisingly given the very low d parameters, the separate 2HTSM analyses for spatial position and frame color were highly similar to those for parameters e^{Position} and e^{Color} in the multidimensional source monitoring model. For detailed information on the originally planned analyses with the 2HTSM, see the preregistration document on the OSF (https://osf.io/a6z4u).

dimension “spatial position” was almost 13% larger in the sleep than in the wake condition. In contrast, there was no statistically significant sleep benefit for the unbound source memory parameter for color e^{Color}, $\Delta G^2(1) = 0.41$, $p = .523$, although there was a descriptive pattern in the predicted direction. Similar to the result for bound source memory, it appears that this unexpected result is a consequence of the low probability of correctly identifying the source (i.e., blue, yellow) of recognized target items on the source dimension “frame color”. In fact, unbound source memory for color ($e^{\text{Color}} \leq .10$) is significantly worse than for spatial position ($e^{\text{Position}} \geq .30$) in both the sleep, $\Delta G^2(1) = 135.92$, $p < .001$, and the wake condition, $\Delta G^2(1) = 46.98$, $p < .001$. Concerning guessing parameters, we found no statistically significant differences between sleep and wake conditions whatsoever, all $\Delta G^2(1) <= 2.19$, $p >= .139$.

As in Experiment 1, we checked the robustness of our results by performing an alternative model-based analysis using Klauer’s (2010) latent-trait model as implemented in TreeBUGS (Heck et al., 2018). Again, as detailed in the Appendix (see Table A2), the estimated group-level means of the hierarchical MPT model closely resembled those reported in Table 4. We conclude that the basic result pattern is robust against using complete versus partial pooling data analysis methods.

Discussion

Replicating Experiment 1 and some of the previous studies (e.g., Mawdsley et al., 2014; van der Helm et al., 2011), we observed a significant sleep benefit in ACSIM-based and model-based source memory for spatial position in Experiment 2. This supports a core prediction of the active systems consolidation hypothesis, namely, that sleep compared to wakefulness should benefit item-context-binding in episodic memory for retention intervals up to 12 hr. However, this core prediction was not confirmed for the second source dimension “frame color”, that is, the corresponding item-context binding parameter, e^{Color}, did not differ significantly between sleep and wake conditions. Moreover, compared to source dimension “spatial position”, source dimension “frame color” exhibited considerably worse source memory in both sleep and wake conditions, suggesting that frame colors were only weakly encoded.

Although some empirical findings suggest that weakly encoded memories profit more from memory consolidation during sleep than stronger encoded memories (e.g., Denis et al., 2020; Drosopoulos et al., 2007), there are other studies showing that stronger memories benefit more (e.g., Tucker & Fishbein, 2008; Schoch et al., 2017). At first glance, these results appear contradictory, they are however in line with the assumption that sleep benefits follow an inverted U-shaped function of memory strength (Stickgold, 2009). According to this account, sleep benefits increase with encoding strength up to a medium level of memory strength before they decrease. In Experiment 2, memory strength likely ranges within the lower limb of the inverted U-shaped function, as indicated by the unbound source memory parameter estimates of the aggregated model-based analysis for spatial position ($e^{\text{Position}} \leq .43$) and frame color ($e^{\text{Color}} \leq .10$). Thus, the consolidation-based sleep benefit in source memory should be larger for spatial position than for frame color. However, for frame color, we observed no significant sleep benefit in source memory. This result can be explained by previous research showing that a minimum level of memory strength at encoding is necessary for the sleep benefit to occur (e.g., Denis et al., 2020; Muehlroth et al., 2020; Rauchs et al., 2011). The insignificant sleep benefit in source memory for frame color may therefore be a consequence of insufficient encoding.

Former studies that used color as source dimension to investigate episodic context-binding enforced intentional learning of item-context associations (e.g., Köster et al., 2017; Wang & Fu, 2009). In contrast, item-context and context-context associations were incidentally learned in our Experiment 2 to create a more realistic setting that resembles everyday source monitoring. Obviously, although we employed an orienting task towards both source dimensions, incidental learning does not seem to support encoding of the frame color context.

In fact, frame color appears to be a less salient context feature than spatial position. For at least two reasons, an item's frame color is less likely to be encoded successfully than its spatial position. First, from an evolutionary perspective, it has often been argued that spatial position is more important for survival and thus receives prioritized processing (Hasher & Zacks, 1979; Nairne et al., 2012; Yin et al., 2019). Second, according to the object-file theory of visual perception (Kahneman et al., 1992; Mitroff et al., 2004, 2005), the spatial position of an object is encoded in a first step by default. Other details such as color are added in a second step that requires more elaborated processing (Chen & Wyble, 2015). Taking these theories into account, it does not come as a surprise that context memory for frame color is considerably worse than context memory for spatial position, irrespective of sleep or wake states during retention.

In terms of item memory, Experiment 2 replicated the mixed results of Experiment 1 and previous research (e.g., Mawdsley et al., 2014; van der Helm et al., 2011). Like in Experiment 1, the observed descriptive result patterns of *d′* and the aggregated as well as hierarchical model-based analyses suggest that sleep compared to wakefulness might benefit item recognition. Again, this pattern was significant only for complete pooling. Hence, the results of Experiment 1 and Experiment 2 might hint to a sleep benefit in item recognition, which is however hard to detect due to its small size.

In sum, Experiment 2 confirms the conclusion drawn from Experiment 1 that sleep benefits binding of salient context features in episodic memory for a 12-hr retention interval as predicted by the active systems consolidation hypothesis. In addition to replicating the results of the first experiment, Experiment 2 also explored whether sleep benefits context-context binding. However, bound source memory for spatial position and frame color did not differ significantly between sleep and wake conditions, most likely as a consequence of floor effects in either condition. Together with the insignificant sleep benefit in source memory for the less salient source dimension “frame color”, this result suggests that a sufficiently high memory strength of item-context and context-context associations at encoding is necessary for the sleep benefit to occur. Whereas spatial position appears to be a context feature that receives sufficient processing during encoding, frame color apparently does not—at least under incidental learning conditions as employed in Experiment 2.

General Discussion

In two experiments, we tested a core assumption of the active systems consolidation hypothesis, namely, that sleep benefits episodic memory context-binding for relatively short retention intervals of up to 12 hr. In contrast to previous research, we made use of MPT models which provide uncontaminated measures of source memory. Both experiments consistently showed a sleep benefit in source memory for spatial position as predicted by the active systems consolidation account. In contrast, the results for item memory were mixed which is in line with previous research.

Using MPT models to decompose source monitoring performance in effects of separate underlying cognitive processes is a powerful alternative to traditional measures. However, there are also caveats that must be considered in MPT modeling. Specifically, by imposing equality constraints on model parameters of source monitoring MPT models, several submodels can be defined, which raises the problem of determining which model to use (Bayen et al., 1996). This problem is usually solved by applying the most parsimonious model that still fits the data (i.e., the model with the smallest number of free parameters). In case of the 2HTSM, the most parsimonious model is Submodel 4 which imposes equality constraints on all item memory parameters, all source memory parameters, and on the guessing parameters *a* and *g*, respectively (Bayen et al., 1996). However, we were forced to relax the equality constraint for the guessing parameters (Submodel 5a), because applying Submodel 4 to the aggregated data of Experiment 1 resulted in misfit (see the Results section of Experiment 1). Alternatively, one could also relax the equality constraint for the source memory parameters instead of the two source guessing parameters, resulting in a data equivalent model (i.e., Submodel 5d of Bayen et al., 1996) that fits the data as well as Submodel 5a. Most important with respect to our research questions, however, both Submodel 5a and 5d showed statistically significant sleep benefits in item memory and in source memory for at least one of the two sources involved.[9] In other words, our substantive conclusions concerning sleep benefits in item

---

[9] Applying Submodel 5d resulted in the same good fit as in case of Submodel 5a, $G^2(2) = 1.78$, $p = .411$. For Submodel 5d, we found a statistically significant sleep benefit in item memory (parameter *D*; sleep: MLE = .65, *SE* = .01; wake: MLE = .60, *SE* = .01), $\Delta G^2(1) = 13.66$, $p < .001$. However, whereas a significant sleep benefit emerged in source memory for source “right” (parameter d_{right}; sleep: MLE = .73, *SE* = .03; wake: MLE = .51, *SE* = .06), $\Delta G^2(1) = 12.28$, $p < .001$, we observed no significant sleep benefit in source memory for source “left”

and source memory would not be severely affected by whether we prefer Submodel 5a or 5d for data analysis.

Fortunately, the more complex design of Experiment 2 with two source dimensions circumvents problems of submodel data equivalency (Bröder & Meiser, 2007). The corresponding multidimensional source monitoring model clearly showed that sleep benefits unbound source memory for both spatial positions (i.e., left and right) to the same degree. Taking this into account, it seems safe to adopt a model with a single source memory parameter (i.e., Submodel 5a) also for Experiment 1. Most importantly, however, all model-based results provide unequivocal evidence for a sleep benefit in episodic memory binding of spatial context features, irrespective of the 2HTSM submodel used.

Aside from the measures used to assess item and source memory, there are further potential moderators that may have contributed to the mixed results in previous research, specifically, the sleep study design employed and the encoding strength of relevant episodic information. As already outlined in the introduction, researchers have used different sleep study designs to investigate sleep benefits in episodic memory context-binding, ranging from split-night designs (e.g., Groch et al., 2015; Rauchs et al., 2004; Sopp et al., 2017), naps (e.g., Köster et al., 2017; van der Helm et al., 2011; Wang & Fu, 2009), to comparisons of natural night sleep with daytime wakefulness (e.g., Lewis et al., 2011; Mawdsley et al., 2014; Wang et al., 2017). All these sleep studies differ in the amount of SWS associated with sleep, a sleep feature assumed to be essential for memory consolidation (e.g., Klinzing et al., 2019; Lewis & Durrant, 2011). In fact, a recent meta-analysis on the sleep benefit in episodic memory showed that sleep benefits tend to be larger for sleep study designs associated

---

(parameter d_{left}; sleep: MLE = .46, *SE* = .08; wake: MLE = .45, *SE* = .06), $\Delta G^2(1) = 0.01$, $p = .919$. Concerning the guessing parameters, we found significantly more “old”-guessing in the wake than in the sleep condition (parameter *b*; sleep: MLE = .20, *SE* = .01; wake: MLE = .25, *SE* = .01), $\Delta G^2(1) = 6.09$, $p = .014$, resembling the result observed with Submodel 5a. However, with respect to a stronger “left” guessing bias after sleep (parameter *a*; sleep: MLE = .64, *SE* = .04; wake: MLE = .47, *SE* = .04), $\Delta G^2(1) = 10.13$, $p = .001$, the result differed from the one achieved with Submodel 5a.

with large amounts of SWS (Berres & Erdfelder, 2021). It is therefore likely that different sleep study designs additionally contributed to the mixed results reported in the literature.

Weak encoding strength may explain why we observed no significant sleep benefit neither in source memory for frame color nor in bound source memory for spatial position and frame color. The fact that both corresponding parameters were at or near floor level suggests that sufficiently high memory strength of item-context and context-context associations at encoding is necessary for the sleep benefit to occur (cf. Denis et al., 2020; Muehlroth et al., 2020; Rauchs et al., 2011). The strength of memory representations is affected by various aspects of the encoding situation, such as presentation time, scope and type of the stimulus material, and encoding instructions. In fact, whereas the current experiments and few others employed incidental learning of item-context associations (e.g., Mawdsley et al., 2014; Wang et al., 2017), most previous studies used intentional learning, in some cases even with the explicit instruction to use mnemonic strategies for context encoding (e.g., Köster et al., 2017; Lewis et al., 2011). Moreover, different source dimensions require different degrees of cognitive effort to be encoded successfully (see the Discussion section of Experiment 2). These aspects, among others, may have affected the strength of the memory representation at encoding, contributing to the mixed results in different sleep studies of source monitoring found thus far.

In line with the active systems consolidation hypothesis, we found convincing evidence for a sleep benefit in source memory using a 12-hr retention interval. Note, however, that sleep benefits in episodic memory can be explained not only by memory consolidation but also by reduced retroactive interference (for a review of consolidation and interference theories, see Berres & Erdfelder, 2021). The contextual binding account (Yonelinas et al., 2019), for example, explains the sleep benefit in terms of a passive effect on memory retrieval. Specifically, retrieval of a target information can be impaired by information learned before or after, provided the content or context of the interfering and the target information resemble each other. In other words, context similarity (i.e., similarity of any aspect of a specific learning situation such as spatial position or color) may foster retroactive interference. During sleep however, new learning is virtually absent. Thus, retroactive interference due to content or context similarity is reduced which in turn should facilitate retrieval of the target information after sleep compared to wakefulness (Yonelinas et al., 2019).

Yet, in terms of source memory, sleep benefits due to reduced retroactive interference appear to play a minor role. According to the memory-system dependent forgetting hypothesis (Hardt et al., 2013; see also Sadeh et al., 2014), interference effects on hippocampal memory representations, such as item-item or item-context associations, should be “minimal” (Hardt et al., 2013, p. 111). As such, the circuit architecture of the hippocampus is assumed to allow efficient pattern separation by assigning orthogonal representations even to highly similar information, thereby diminishing overlapping neuronal populations and thus interference. By contrast, memory representations like item memory, that are linked to extrahippocampal regions, are represented by overlapping neuronal populations. As a consequence, these memories should be very susceptible to interference (Hardt et al., 2013). Indeed, supporting this theory, Kuhlmann et al. (2021) investigated forgetting over short, interference-filled lags in three experiments and found pronounced interference-based forgetting in item memory compared to item-item and item-context associative memory (see also Sadeh et al., 2014).

Taking the memory-system dependent forgetting hypothesis into account when considering underlying processes of the sleep benefit in source memory, two assumptions can be made: First, because for source memory interference effects should be minimal, it can be assumed that sleep benefits in item-item and item-context associative memory depend more on memory consolidation than on reduced retroactive interference. Therefore, we interpret our results for source memory in terms of the active systems consolidation hypothesis, although additional sleep benefits on memory retrieval as predicted by the contextual binding account cannot be ruled out completely.

Second, for item memory, which should be very susceptible to interference, it can be assumed that sleep benefits are more heavily based on retrieval advantages due to reduced retroactive interference compared to source memory. This assumption can also offer a possible explanation for the mixed evidence of sleep benefits in item memory. To reiterate, the results of Experiment 1 and Experiment 2 might hint to a small sleep benefit in item recognition. The small size of the effect might be due to the fact, that we assessed item memory with a recognition task. Recent meta-analyses suggest that the sleep benefit is moderated by the retrieval procedure (e.g., Berres & Erdfelder, 2021; Newbury & Monaghan, 2019). As such, free recall relies more heavily on memory retrieval than cued recall, and cued recall more than recognition. Thus, sleep benefits due to reduced retroactive interference should be largest in free recall, followed by cued recall, and lastly recognition tasks (cf. Dyne et al., 1990; McKinney, 1935; Postman, 1952). This might explain why findings regarding sleep benefits in item memory are mixed.

Apart from the processes underlying the sleep benefit in item and source memory, several open questions remain for future research. First, probably because bound source memory parameter *d* was at floor level (Experiment 2), no significant sleep benefit in context-context binding emerged. Future studies should therefore ensure that context-context associations are encoded with sufficient memory strength to allow for a rigorous test of the hypothesis that sleep benefits bound source memory. This may, however, require switching to intentional learning instructions and associative encoding strategies and thus a very specific type of encoding only.

Second, in the current experiments, we investigated the sleep benefit in source memory using a 12 hr retention interval. However, the precise time course of memory consolidation during sleep is not yet well understood (Dudai, 2004, 2012; Dudai et al., 2015; Klinzing et al., 2019; Lewis & Durrant, 2011; Pöhlchen & Schönauer, 2020; Stickgold, 2005; Stickgold & Walker, 2007). Further studies using standardized designs with different retention intervals are therefore necessary to explore the minimum and maximum retention interval for which sleep compared to wakefulness benefits source memory.

Third, context-recollection in episodic memory is often not only assessed with source memory measures but also with remember-know judgments to capture subjective retrieval experiences with respect to conscious recollection and familiarity in addition (for a review, see Inostroza & Born, 2013). In the current experiments, we used source memory measures only because they reflect hippocampus-dependent memories more directly than remember-know judgments. Nevertheless, extending model-based analyses to include remember-know judgments (see, for example, Meiser, 2014) may provide more fine-graded insights into sleep-dependent benefits in episodic memory context-binding. In this context it is particularly interesting to investigate less hippocampus-dependent memories in future research, as recent studies suggest that memories which likely do not require the hippocampus during encoding may depend on it for consolidation during sleep (e.g., Sawangjit et al., 2018; Schapiro et al., 2019).

Conclusion

Our two experiments consistently show that sleep benefits source memory, provided that relevant context features—such as the spatial position of an item—are sufficiently salient and thus well encoded. These results are in line with the prediction of the active systems consolidation hypothesis that sleep benefits item-context-binding for 12-hr retention intervals. In addition, our findings call attention to potential moderators that may explain the mixed results in previous research, such as level of analysis employed or the encoding strength of source dimensions that prevents sleep benefits in context-context bindings when at least one source dimension is insufficiently encoded (see Experiment 2). In sum, the present research adds to the growing empirical evidence that memory consolidation as described by the active systems consolidation hypothesis is one of the key neurocognitive processes that contributes to the sleep benefit in episodic memory.

References

Abel, M., & Bäuml, K.-H. T. (2012). Retrieval-induced forgetting, delay, and sleep. *Memory, 20*(5), 420–428. https://doi.org/10.1080/09658211.2012.671832

Abel, M., & Bäuml, K.-H. T. (2013a). Adaptive memory: The influence of sleep and wake delay on the survival-processing effect. *Journal of Cognitive Psychology, 25*(8), 917–924. https://doi.org/10.1080/20445911.2013.825621

Abel, M., & Bäuml, K.-H. T. (2013b). Sleep can eliminate list-method directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(3), 946–952. https://doi.org/10.1037/a0030529

Abel, M., & Bäuml, K.-H. T. (2014). Sleep can reduce proactive interference. *Memory, 22*(4), 332–339. https://doi.org/10.1080/09658211.2013.785570

Ambrosini, M. V., & Giuditta, A. (2001). Learning and sleep: The sequential hypothesis. Sleep *Medicine Reviews, 5*(6), 477–490. https://doi.org/10.1053/smrv.2001.0180

Arnold, N. R., Heck, D. W., Bröder, A., Meiser, T., & Boywitt, C. D. (2019). Testing hypotheses about binding in context memory with a hierarchical multinomial modeling approach: A preregistered study. *Experimental Psychology, 66*(3), 239–251. https://doi.org/10.1027/1618-3169/a000442

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review, 97*(4), 548–564. https://doi.org/10.1037/0033-295X.97.4.548

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*(1), 57–86. https://doi.org/10.3758/BF03210812

Batchelder, W. H., Riefer, D. M., & Hu, X. (1994). Measuring memory factors in source monitoring: Replay to Kinchla. *Psychological Review, 101*(1), 172–176. https://doi.org/10.1037/0033-295X.101.1.172

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(1), 197–215. https://doi.org/10.1037/0278-7393.22.1.197

Bäuml, K.-H. T., Holterman, C., & Abel, M. (2014). Sleep can reduce the testing effect: It enhances recall of restudied items but can leave recall of retrieved items unaffected. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(6), 1568–1581. https://doi.org/10.1037/xlm0000025

Bell, R., Mieth, L., & Buchner, A. (2017). Emotional memory: No source memory without old-new recognition. *Emotion, 17*(1), 120–130. https://doi.org/10.1037/emo0000211

Berres, S., & Erdfelder, E. (2021). The sleep benefit in episodic memory: An integrative review and a meta-analysis. *Psychological Bulletin, 147*(12), 1309–1353. https://doi.org/10.1037/bul0000350

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

Boywitt, C. D., & Meiser, T. (2012). The role of attention for context-context binding of intrinsic and extrinsic features. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(4), 1099–1107. https://doi.org/10.1037/a0026988

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLOS ONE, 5*(5), Article e10773. https://doi.org/10.1371/journal.pone.0010773

Bröder, A., & Meiser, T. (2007). Measuring source memory. *Zeitschrift für Psychologie / Journal of Psychology, 215*(1), 52–60. https://doi.org/10.1027/0044-3409.215.1.52

Buzsáki, G. (1998). Memory consolidation during sleep: A neurophysiological perspective. *Journal of Sleep Research*, *7*(S1), 17–23. https://doi.org/10.1046/j.1365-2869.7.s1.3.x

Chen, H., & Wyble, B. (2015). The location but not the attributes of visual cues are automatically encoded into working memory. *Vision Research, 107*, 76–85. https://doi.org/10.1016/j.visres.2014.11.010

Cirelli, C., & Tononi, G. (2015). Sleep and synaptic homeostasis. *Sleep, 38*(1), 161–162. https://doi.org/10.5665/sleep.4348

Denis, D., Schapiro, A. C., Poskanzer, C., Bursal, V., Charon, L., Morgan, A., & Stickgold, R. (2020). The role of item exposure and visualization success in the consolidation of memories across wake and sleep. *Learning & Memory, 27*, 451–456. http://www.learnmem.org/cgi/doi/10.1101/lm.051383.120

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

Diekelmann, S., Wilhelm, I., & Born, J. (2009). The whats and whens of sleep-dependent memory consolidation. *Sleep Medicine Reviews*, *13*(5), 309–321. https://doi.org/10.1016/j.smrv.2008.08.002

Drosopoulos, S., Schulze, C., Fischer, S., & Born, J. (2007). Sleep's function in the spontaneous recovery and consolidation of memories. *Journal of Experimental Psychology: General, 136*(2), 169–183. https://doi.org/10.1037/0096-3445.136.2.169

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annual Review of Psychology*, *55*, 51–86. https://doi.org/10.1146/annurev.psych.55.090902.142050

Dudai, Y. (2012). The restless engram: Consolidations never end. *Annual Review of Neuroscience*, *35*, 227–247. https://doi.org/10.1146/annurev-neuro-062111-150500

Dudai, Y., Karni, A., & Born, J. (2015). The consolidation and transformation of memory. *Neuron*, *88*(1), 20–32. https://doi.org/10.1016/j.neuron.2015.09.004

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology, 71*(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261

Dyne, A. M., Humphreys, M. S., Bain, J. D., & Pike, R. (1990). Associative interference effects in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(5), 813–824. https://doi.org/10.1037/0278-7393.16.5.813

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience, 30*, 123–152. https://doi.org/10.1146/annurev.neuro.30.051606.094328

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie*, *217*(3), 108–124. https://doi.org/10.1027/0044-3409.217.3.108

Erdfelder, E., Berres, S., Quevedo Pütter, J., & Küpper-Tetzel, C. E. (2022). *Why does sleep improve episodic memory? An encoding-maintenance-retrieval analysis* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Feld, G. B., & Born, J. (2017). Sculpting memory during sleep: Concurrent consolidation and forgetting. *Current Opinion in Neurobiology*, *44*, 20–27. https://doi.org/10.1016/j.conb.2017.02.012

Fenn, K. M., & Hambrick, D. Z. (2013). What drives sleep-dependent memory consolidation: Greater gain or less loss? *Psychonomic Bulletin & Review, 20*, 501–506. https://doi.org/10.3758/s13423-012-0366-z

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Giuditta, A. (2014). Sleep memory processing: The sequential hypothesis. Frontiers in Systems *Neuroscience, 8*, Article 219. https://doi.org/10.3389/fnsys.2014.00219

Giuditta, A., Ambrosini, M. V., Montagnese, P., Mandile, P., Cotugno, M., Zucconi, G. G., & Vescia, S. (1995). The sequential hypothesis of the function of sleep. *Behavioural Brain Research, 69*(1–2), 157–166. https://doi.org/10.1016/0166-4328(95)00012-I

Groch, S., Zinke, K., Wilhelm, I., & Born, J. (2015). Dissociating the contributions of slow-wave sleep and rapid eye movement sleep to emotional item and source memory. *Neurobiology of Learning and Memory, 122*, 122–130. https://doi.org/10.1016/j.nlm.2014.08.013

Hardt, O., Nader, K., & Nadel, L. (2013). Decay happens: The role of active forgetting in memory. *Trends in Cognitive Sciences, 17*(3), 111–120. https://dx.doi.org/10.1016/j.tics.2013.01.001

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*(3), 356–388. https://doi.org/10.1037/0096-3445.108.3.356

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d'. *Behavior Research Methods, Instruments, & Computers, 27*(1), 46–51. https://doi.org/10.3758/BF03203619

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods, 50*, 264–284. https://doi.org/10.3758/s13428-017-0869-7

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). lab.js: A free, open, online study builder. *Behavior Research Methods, 54*, 556–573. https://doi.org/10.3758/s13428-019-01283-5

Inostroza, M., & Born, J. (2013). Sleep for preserving and transforming episodic memory. *Annual Review of Neuroscience*, *36*, 79–102. https://doi.org/10.1146/annurev-neuro-062012-170429

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3–28. https://doi.org/10.1037/0033-2909.114.1.3

Jurewicz, K., Cordi, M. J., Staudigl, T., & Rasch, B. (2016). No evidence for memory decontextualization across one night of sleep. *Frontiers in Human Neuroscience*, *10*, Article 7. https://www.frontiersin.org/article/10.3389/fnhum.2016.00007

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*(2), 175–219. https://doi.org/10.1016/0010-0285(92)90007-O

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika, 75*(1), 70–98. https://doi.org/10.1007/s11336-009-9141-0

Klinzing, J. G., Niethard, N., & Born, J. (2019). Mechanisms of systems memory consolidation during sleep. *Nature Neuroscience*, *22*(10), 1598–1610. https://doi.org/10.1038/s41593-019-0467-3

Köster, M., Finger, H., Kater, M.-J., Schenk, C., & Gruber, T. (2017). Neuronal oscillations indicate sleep-dependent changes in the cortical memory trace. *Journal of Cognitive Neuroscience, 29*(4), 698–707. https://doi.org/10.1162/jocn_a_01071

Kuhlmann, B. G., Bayen, U. J., Meuser, K., & Kornadt, A. E. (2016). The impact of age stereotypes on source monitoring in younger and older adults. *Psychology and Aging, 31*(8), 875–889. https://doi.org/10.1037/pag0000140

Kuhlmann, B. G., Brubaker, M. S., Pfeiffer, T., & Naveh-Benjamin, M. (2021). Longer resistance of associative versus item memory to interference-based forgetting, even in older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(3), 422–438. https://dx.doi.org/10.1037/xlm0000963

Leiner, D. J. (2020). *SoSci Survey* (Version 3.2.12) [Computer software]. https://www.soscisurvey.de

Lewis, P. A., Cairney, S., Manning, L., & Critchley, H. D. (2011). The impact of overnight consolidation upon memory for emotional and neutral encoding contexts. *Neuropsychologia*, *49*(9), 2619–2629. https://doi.org/10.1016/j.neuropsychologia.2011.05.009

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

Lutz, N. D., Diekelmann, S., Hinse-Stern, P., Born, J., & Rauss, K. (2017). Sleep supports the slow abstraction of gist from visual perceptual memories. *Scientific Reports*, *7*, Article 42950. https://doi.org/10.1038/srep42950

Mawdsley, M., Grasby, K., & Talk, A. (2014). The effect of sleep on item recognition and source memory recollection among shift-workers and permanent day-workers. *Journal of Sleep Research*, *23*(5), 538–544. https://doi.org/10.1111/jsr.12149

Mayes, A., Montaldi, D., & Migo, E. (2007). Associative memory and the medial temporal lobes. *Trends in Cognitive Sciences, 11*(3), 126–135. https://doi.org/10.1016/j.tics.2006.12.003

McKinney, F. (1935). Retroactive inhibition and recognition memory. *Journal of Experimental Psychology, 18*(5), 585–598. https://doi.org/10.1037/h0054474

Meiser, T. (2014). Analyzing stochastic dependence of cognitive processes in multidimensional source recognition. *Experimental Psychology, 61*(5), 402–415. https://doi.org/10.1027/1618-3169/a000261

Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 116–137. https://doi.org/10.1037/0278-7393.28.1.116

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neuronal mechanisms of source memory? *Psychological Bulletin, 135*(4), 638–677. https://doi.org/10.1037/a0015849

Mitroff, S. R., Scholl, B. J., & Wynn, K. (2004). Divide and conquer: How object files adapt when a persisting object splits into two. *Psychological Science, 15*(6), 420–425. https://doi.org/10.1111/j.0956-7976.2004.00695.x

Mitroff, S. R., Scholl, B. J., & Wynn, K. (2005). The relationship between object files and conscious perception. *Cognition, 96*(1), 67–92. https://doi.org/10.1016/j.cognition.2004.03.008

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods, 42*(1), 42–54. https://doi.org/10.3758/BRM.42.1.42

Muehlroth, B. E., Rasch, B., & Werkle-Bergner, M. (2020). Episodic memory consolidation during sleep in healthy aging. *Sleep Medicine Reviews, 52*, Article 101304. https://doi.org/10.1016/j.smrv.2020.101304

Murnane, K, & Bayen, U. J. (1996). An evaluation of empirical measures of source identification. *Memory & Cognition, 24*(4), 417–428. https://doi.org/10.3758/BF03200931

Nairne, J. S., VanArsdall, J. E., Pandeirada, J. N. S., & Blunt, J. R. (2012). Adaptive memory: Enhanced location memory after survival processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(2), 495–501. https://doi.org/10.1037/a0025728

Newbury, C. R., & Monaghan, P. (2019). When does sleep affect veridical and false memory consolidation? A meta-analysis. *Psychonomic Bulletin & Review, 26*, 387–400. https://doi.org/10.3758/s13423-018-1528-4

O'Neill, J., Pleydell-Bouverie, B., Dupret, D., & Csicsvari, J. (2010). Play it again: Reactivation of waking experience and memory. *Trends in Neurosciences*, *33*(5), 220–229. https://doi.org/10.1016/j.tins.2010.01.006

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance, 17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Pfeiffer, B. E. (2020). The content of hippocampal "replay". *Hippocampus*, *30*(1), 6–18. https://doi.org/10.1002/hipo.22824

Postman, L. (1952). Retroactive inhibition in recall and recognition. *Journal of Experimental Psychology, 44*(3), 165–169. https://doi.org/10.1037/h0057388

Pöhlchen, D., & Schönauer, M. (2020). Sleep-dependent memory consolidation in the light of rapid neocortical plasticity. *Current Opinion in Behavioral Sciences, 33*, 118–125. https://doi.org/10.1016/j.cobeha.2020.02.001

Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Reviews*, *93*(2), 681–766. https://doi.org/10.1152/physrev.00032.2012

Rauchs, G., Bertran, F., Guillery-Girard, B., Desgranges, B., Kerrouche, N., Denise, P., Foret, J., & Eustache, F. (2004). Consolidation of strictly episodic memories mainly requires rapid eye movement sleep. *Sleep*, *27*(3), 395–401. https://doi.org/10.1093/sleep/27.3.395

Rauchs, G., Feyers, D., Landeau, B., Bastin, C., Luxen, A., Maquet, P., & Collette, F. (2011). Sleep contributes to the strengthening of some memories over others, depending on hippocampal activity at learning. *Journal of Neuroscience*, *31*(7), 2563–2568. https://doi.org/10.1523/JNEUROSCI.3972-10.2011

Sadeh, T., Ozubko, J. D., Winocur, G., & Moscovitch, M. (2014). How we forget may depend on how we remember. Trends in Cognitive Sciences, 18(1), 26–36. https://dx.doi.org/10.1016/j.tics.2013.10.008

Sawangjit, A., Oyanedel, C. N., Niethard, N., Salazar, C., Born, J., & Inostroza, M. (2018). The hippocampus is crucial for forming non-hippocampal long-term memory during sleep. *Nature, 564,* 109–113. https://doi.org/10.1038/s41586-018-0716-8

Schapiro, A. C., Reid, A. G., Morgan, A., Manoach, D. S., Verfaellie, M., & Stickgold, R. (2019). The hippocampus is necessary for the consolidation of a task that does not require the hippocampus for initial learning. *Hippocampus, 29*(11), 1091–1100. https://doi.org/10.1002/hipo.23101

Schmidt, O., Erdfelder, E., & Heck, D. W. (2023). How to develop, test, and extend multinomial processing tree models: A tutorial. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000561

Schoch, S. F., Cordi, M. J., & Rasch, B. (2017). Modulating influences of memory strength and sensitivity of the retrieval test on the detectability of the sleep consolidation effect. *Neurobiology of Learning and Memory, 145*, 181–189. https://doi.org/10.1016/j.nlm.2017.10.009

Sopp, M. R., Michael, T., Weeß, H.-G., & Mecklinger, A. (2017). Remembering specific features of emotional events across time: The role of REM sleep and prefrontal theta oscillations. *Cognitive,* *Affective, & Behavioral Neuroscience, 17*, 1186–1209. https://doi.org/10.3758/s13415-017-0542-8

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137–149. https://doi.org/10.3758/BF03207704

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, *437*, 1272–1278. https://doi.org/10.1038/nature04286

Stickgold, R. (2009). How do I remember? Let me count the ways. *Sleep Medicine Reviews, 13*(5), 305–308. https://doi.org/10.1016/j.smrv.2009.05.004

Stickgold, R., & Walker, M. P. (2007). Sleep-dependent memory consolidation and reconsolidation. *Sleep Medicine*, *8*(4), 331–343. https://doi.org/10.1016/j.sleep.2007.03.011

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

Tononi, G., & Cirelli, C. (2006). Sleep function and synaptic homeostasis. *Sleep Medicine Reviews, 10*(1), 49–62. https://doi.org/10.1016/j.smrv.2005.05.002

Tononi, G., & Cirelli, C. (2014). Sleep and the price of plasticity: From synaptic and cellular homeostasis to memory consolidation and integration. *Neuron, 81*(1), 12–34. https://doi.org/10.1016/j.neuron.2013.12.025

Tononi, G., & Cirelli, C. (2020). Sleep and synaptic down-selection. *European Journal of Neuroscience, 51*(1), 413–421. https://doi.org/10.1111/ejn.14335

Tucker, M. A., & Fishbein, W. (2008). Enhancement of declarative memory performance following a daytime nap is contingent on strength of initial task acquisition. *Sleep, 31*(2), 197–203. https://doi.org/10.1093/sleep/31.2.197

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53*, 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114 van der Helm, E., Gujar, N., Nishida, M., & Walker, M. P. (2011). Sleep-dependent facilitation of episodic memory details. *PLOS ONE*, *6*(11), Article e27421. https://doi.org/10.1371/journal.pone.0027421

Wang, B., & Fu, X.-l. (2009). Gender difference in the effect of daytime sleep on declarative memory for pictures. *Journal of Zhejiang University SCIENCE B*, *10*(7), 536–546. https://doi.org/10.1631/jzus.B0820384

Wang, J.-Y., Weber, F. D., Zinke, K., Noack, H., & Born, J. (2017). Effects of sleep on word pair memory in children—Separating item and source memory aspects. *Frontiers in Psychology*, *8*, Article 1533. https://doi.org/10.3389/fpsyg.2017.01533

Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, *265*(5172), 676–679. https://doi.org/10.1126/science.8036517

Yin, X., Ma, Y., Xu, X., & Yang, H. (2019). The effect of self-referencing on memory for different kinds of source information. *Memory, 27*(4), 519–527. https://doi.org/10.1080/09658211.2018.1532009

Yonelinas, A. P., Ranganath, C., Ekstrom, A. D., & Wiltgen, B. J. (2019). A contextual binding theory of episodic memory: Systems consolidation reconsidered. *Nature Reviews Neuroscience, 20*, 364–375. https://doi.org/10.1038/s41583-019-0150-4

Table 1

Results of Item-Memory Analysis in Experiment 1

Dependent variable	Wake		Sleep			
	M	*SE*	*M*	*SE*	*t*(130)	*p*
Hit rate	0.70	0.02	0.72	0.02	0.65	.517
False-alarm rate	0.10	0.01	0.07	0.01	1.62	.107
Sensitivity index *d′*	2.05	0.09	2.27	0.08	1.35	.181
Response bias *c*	0.41	0.03	0.43	0.03	0.31	.759

Note. Means and standard errors of the mean are shown for the wake (*n* = 65) and sleep condition (*n* = 67), as well as the results of two-tailed *t* tests comparing the two independent groups.

Table 2

Aggregated Parameter Estimates of the Two-High-Threshold Multinomial Model of Source Monitoring (2HTSM) for Experiment 1

Parameter	Wake			Sleep		
	MLE	*SE*	95% CI	MLE	*SE*	95% CI
D	.60	.01	[.58, .62]	.65	.01	[.64, .67]
d	.48	.02	[.44, .52]	.63	.02	[.60, .67]
a	.44	.02	[.39, .48]	.48	.03	[.43, .53]
b	.25	.01	[.22, .27]	.20	.01	[.17, .23]
g	.47	.04	[.40, .54]	.64	.04	[.57, .72]

Note. For the aggregated model-based analysis, maximum likelihood estimates (MLE), standard errors (*SE*), and 95% confidence intervals (CI) are reported. *D* = probability of correctly identifying a target item as “old” and a distractor item as “new”; *d* = probability of correctly identifying the source of a target item; *a* = probability of guessing that a correctly identified target item is from source “left”; *b* = probability of guessing that an item is “old”; *g* = probability of guessing that an unrecognized item is from source “left” if it was guessed to be “old”.

Table 3

Results of Item-Memory Analysis in Experiment 2

Dependent variable	Wake		Sleep			
	M	*SE*	*M*	*SE*	*t*(132)	*p*
Hit rate	0.47	0.02	0.52	0.02	1.29	.199
False-alarm rate	0.14	0.01	0.13	0.01	0.55	.581
Sensitivity index *d′*	1.16	0.07	1.32	0.07	1.25	.212
Response bias *c*	0.66	0.04	0.60	0.04	0.72	.474

Note. Means and standard errors of the mean are shown for the wake (*n* = 62) and sleep condition (*n* = 72), as well as the results of two-tailed *t* tests comparing the two independent groups.

Table 4

Aggregated Parameter Estimates of the Multinomial Model of Multidimensional Source Monitoring for Experiment 2

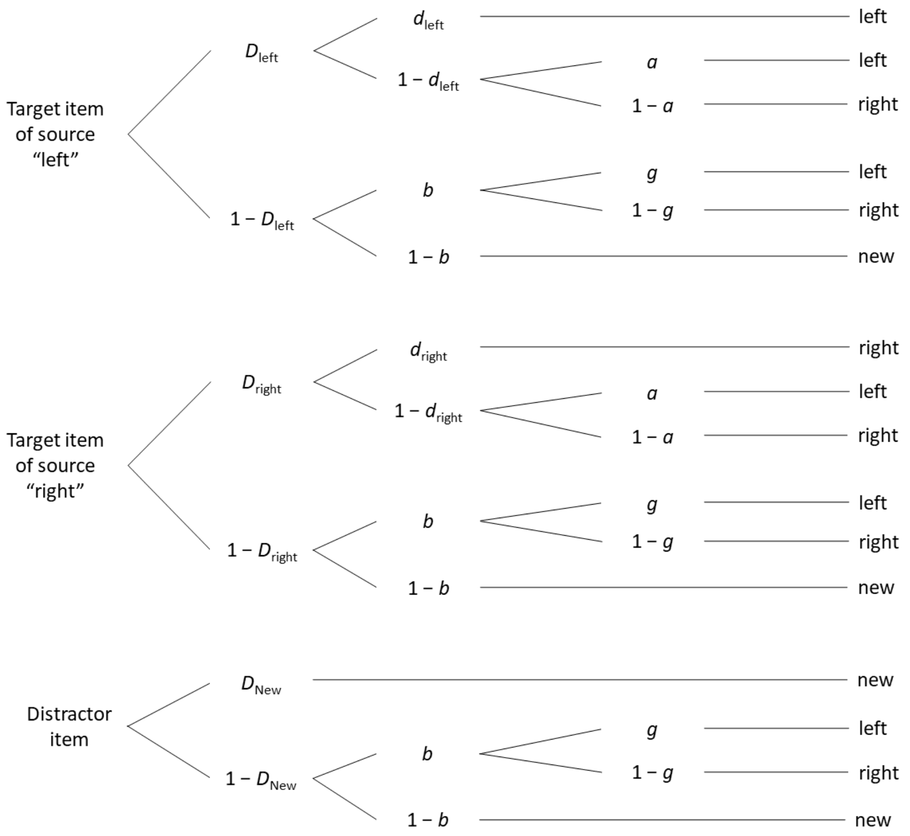
Parameter	Wake			Sleep		
	MLE	*SE*	95% CI	MLE	*SE*	95% CI
D	.33	.01	[.32, .35]	.40	.01	[.38, .41]
d	.04	.04	[.00, .11][a]	.04	.04	[.00, .11][a]
e^{Position}	.30	.04	[.23, .37]	.43	.03	[.37, .49]
e^{Color}	.07	.04	[.00, .15][a]	.10	.04	[.02, .18]
a^{Position}	.48	.02	[.44, .52]	.43	.02	[.39, .47]
$a_{\mid\text{left}}{}^{\text{Color}}$	.51	.02	[.46, .55]	.50	.02	[.46, .53]
$a_{\mid\text{right}}{}^{\text{Color}}$	.61	.03	[.54, .67]	.57	.03	[.51, .63]
b	.21	.01	[.19, .22]	.21	.01	[.19, .22]
g^{Position}	.51	.02	[.46, .55]	.55	.02	[.50, .59]
$g_{\mid\text{left}}{}^{\text{Color}}$	.49	.03	[.43, .55]	.53	.03	[.48, .58]
$g_{\mid\text{right}}{}^{\text{Color}}$	.47	.03	[.40, .53]	.52	.03	[.45, .58]

Note. For the aggregated model-based analysis, maximum likelihood estimates (MLE), standard errors (*SE*), and 95% confidence intervals (CI) are reported. D = probability of correctly identifying a target item as "old" and a distractor item as "new"; d = probability of correctly identifying the source combination of a target item; e^{Position} = probability of correctly identifying the source (i.e., left, right) on source dimension "spatial position" if the target item was correctly identified as "old"; e^{Color} = probability of correctly identifying the source (i.e., blue, yellow) on source dimension "frame color" if the target item was correctly identified as "old"; a^{Position} = probability of guessing ''left" on source dimension "spatial position" if the target item was correctly identified as "old"; $a_{\mid\text{left}}{}^{\text{Color}}$ = probability of guessing "blue" on source dimension "frame color" if the target item was correctly identified as "old" and assigned to source "left"; $a_{\mid\text{right}}{}^{\text{Color}}$ = probability of guessing "blue" on source dimension “frame color” if the target item was correctly identified as “old” and assigned to source “right”; *b* = probability of guessing that an item is “old”; g^{Position} = probability of guessing “left” on source dimension “spatial position” if the unrecognized item was guessed to be “old”; $g_{|\text{left}}^{\text{Color}}$ = probability of guessing “blue” on source dimension “frame color” if the unrecognized item was guessed to be “old” and assigned to source “left”; $g_{|\text{right}}^{\text{Color}}$ = probability of guessing “blue” on source dimension “frame color” if the unrecognized item was guessed to be “old” and assigned to source “right”.

[a] Asymptotic confidence interval boundaries with values below .00 or above 1.00 were set to 0 and 1, respectively, because the parameter space limits the range of admissible values to [.00, 1.00].

Figure 1

Two-High-Threshold Multinomial Model of Source Monitoring (2HTSM) Adapted to the Spatial Position Source Manipulation Used in Experiment 1

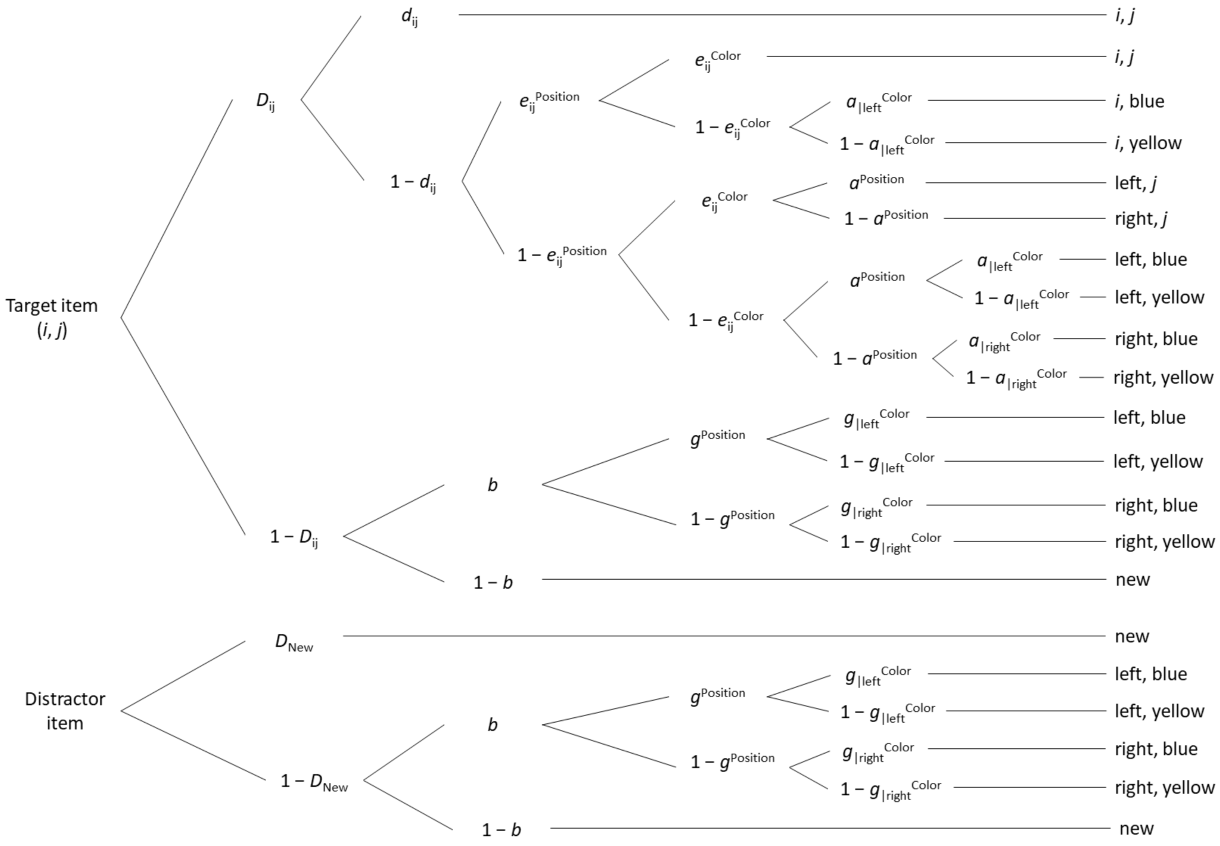


Note. D_{left} = probability of correctly identifying a target item in source "left" as "old"; D_{right} = probability of correctly identifying a target item in source "right" as "old"; D_{New} = probability of correctly identifying a distractor item as "new"; d_{left} = probability of correctly identifying the source of a target item in source "left"; d_{right} = probability of correctly identifying the source of a target item in source "right"; a = probability of guessing that a correctly identified target item is from source "left"; b = probability of guessing that an item is "old"; g = probability of guessing that an unrecognized item is from source "left" if it was guessed to be "old". Adapted from "Source Discrimination, Item Detection, and Multinomial Models of Source Monitoring", by U. J. Bayen, K. Murnane, and E. Erdfelder, 1996, *Journal of Experimental Psychology: Learning, Memory, and*

Cognition, 22(1), p. 202 (https://doi.org/10.1037/0278-7393.22.1.197). Copyright 1996 by the American Psychological Association.

Figure 2

Multinomial Model of Multidimensional Source Monitoring Used in Experiment 2



Note. D_{ij} = probability of correctly recognizing a target item in sources *i* (i.e., “left” or “right” on source dimension “spatial position”) and *j* (i.e., “blue” or “yellow” on source dimension “frame color”) of both source dimensions; D_{New} = probability of correctly identifying a distractor item as “new”; d_{ij} = probability of correctly identifying the source combination *i*, *j* of a recognized item (i.e., “left and blue”, “left and yellow”, “right and blue”, or “right and yellow”, respectively); $e_{ij}^{Position}$ = probability of correctly identifying the source (i.e., left, right) on source dimension “spatial position” of a recognized item; e_{ij}^{Color} = probability of correctly identifying the source (i.e., blue, yellow) on source dimension “frame color” of a recognized item; $a^{Position}$ = probability of guessing ‘‘left” on source dimension “spatial position” of a recognized item; $a_{|left}^{Color}$ = probability of guessing “blue” on source dimension “frame color” if the target item was correctly identified as “old” and assigned to source “left”; $a_{|right}^{Color}$ = probability of guessing “blue” on source dimension “frame color” if the target item was correctly identified as “old” and assigned to source “right”; *b* = probability of guessing that an unrecognized item is “old”; $g^{Position}$ = probability of guessing “left” on source dimension “spatial position” if the unrecognized item was guessed to be “old”; $g_{|left}^{Color}$ = probability of guessing “blue” on source dimension “frame color” if the unrecognized item was guessed to be “old” and assigned to source “left”; $g_{|right}^{Color}$ = probability of guessing “blue” on source dimension “frame color” if the unrecognized item was guessed to be “old” and assigned to source “right”. Adapted from “Analyzing Stochastic Dependence of Cognitive Processes in Multidimensional Source Recognition”, by T. Meiser, 2014, *Experimental Psychology, 61*(5), p. 408 (https://doi.org/10.1027/1618-3169/a000261). Copyright 2014 by Hogrefe Publishing.

Appendix

Results of the Bayesian-Hierarchical Multinomial Processing Tree (MPT) Model Analyses

Table A1

Hierarchical Bayesian Parameter Estimates of the Latent-Trait Version of the Two-High-Threshold Multinomial Model of Source Monitoring (2HTSM) for Experiment 1

Parameter	Wake			Sleep		
	M	*SD*	95% BCI	*M*	*SD*	95% BCI
D	.62	.03	[.55, .68]	.68	.03	[.62, .73]
d	.42	.05	[.33, .51]	.62	.04	[.54, .69]
a	.44	.03	[.37, .51]	.48	.03	[.42, .53]
b	.18	.03	[.13, .24]	.17	.03	[.11, .23]
g	.49	.05	[.40, .58]	.70	.08	[.53, .86]

Note. For the latent-trait model (Klauer, 2010), posterior means (*M*), posterior standard deviations (*SD*), and 95% Bayesian credibility intervals (BCI) of the probability transformed group-level parameters as estimated with TreeBUGS (Heck et al., 2018) are reported. There was good MCMC chain convergence ($\hat{R} < 1.05$) and model fit (wake condition: $p_{T1} = .39$, $p_{T2} = .50$; sleep condition: $p_{T1} = .58$, $p_{T2} = .41$). *D* = probability of correctly identifying a target item as "old" and a distractor item as "new"; *d* = probability of correctly identifying the target item source; *a* = probability of guessing that a correctly identified target item is from source "left"; *b* = probability of guessing that an item is "old"; *g* = probability of guessing that an unrecognized item is from source "left" if it was guessed to be "old"; $\hat{R}$ = potential scale reduction factor (Gelman & Rubin, 1992); p_{T1} = posterior predictive *p*-value for the mean; p_{T2} = posterior predictive *p*-value for the covariance.

Table A2

Hierarchical Bayesian Parameter Estimates of the Latent-Trait Version of the Multinomial Model of Multidimensional Source Monitoring for Experiment 2

Parameter	Wake			Sleep		
	M	*SD*	95% BCI	*M*	*SD*	95% BCI
D	.29	.03	[.23, .36]	.36	.03	[.31, .42]
d	.01	.01	[.00, .03]	.01	.01	[.00, .04]
e^{Position}	.21	.05	[.10, .31]	.34	.06	[.22, .46]
e^{Color}	.01	.01	[.00, .04]	.01	.01	[.00, .03]
a^{Position}	.44	.03	[.38, .50]	.45	.03	[.40, .51]
$a_{\mid\text{left}}^{\text{Color}}$	.54	.04	[.46, .61]	.52	.02	[.47, .56]
$a_{\mid\text{right}}^{\text{Color}}$	.63	.05	[.54, .75]	.58	.05	[.48, .68]
b	.17	.02	[.13, .22]	.19	.02	[.15, .23]
g^{Position}	.53	.03	[.47, .58]	.53	.03	[.47, .59]
$g_{\mid\text{left}}^{\text{Color}}$	.46	.04	[.39, .53]	.48	.04	[.40, .56]
$g_{\mid\text{right}}^{\text{Color}}$	.47	.03	[.40, .54]	.51	.04	[.44, .59]

Note. For the latent-trait model (Klauer, 2010), posterior means (*M*), posterior standard deviations (*SD*), and 95% Bayesian credibility intervals (BCI) of the probability transformed group-level parameters as estimated with TreeBUGS (Heck et al., 2018) are reported. There was good MCMC chain convergence ($\hat{R} < 1.05$) and model fit (wake condition: $p_{T1} = .52$, $p_{T2} = .42$; sleep condition: $p_{T1} = .05$, $p_{T2} = .48$). D = probability of correctly identifying a target item as "old" and a distractor item as "new"; d = probability of correctly identifying the source combination of a target item; e^{Position} = probability of correctly identifying the source (i.e., left, right) on source dimension "spatial position" independent from source dimension "frame color" if the target item was correctly identified as "old"; e^{Color} = probability of correctly identifying the source (i.e., blue, yellow) on source dimension "frame color" independent from source dimension "spatial position" if the target item was correctly identified as “old”; a^{Position} = probability of guessing ‘‘left” on source dimension “spatial position” if the target item was correctly identified as “old”; $a_{|\text{left}}^{\text{Color}}$ = probability of guessing “blue” on source dimension “frame color” if the target item was correctly identified as “old” and assigned to source “left”; $a_{|\text{right}}^{\text{Color}}$ = probability of guessing “blue” on source dimension “frame color” if the target item was correctly identified as “old” and assigned to source “right”; b = probability of guessing that an item is “old”; g^{Position} = probability of guessing “left” on source dimension “spatial position” if the unrecognized item was guessed to be “old”; $g_{|\text{left}}^{\text{Color}}$ = probability of guessing “blue” on source dimension “frame color” if the unrecognized item was guessed to be “old” and assigned to source “left”; $g_{|\text{right}}^{\text{Color}}$ = probability of guessing “blue” on source dimension “frame color” if the unrecognized item was guessed to be “old” and assigned to source “right”; $\hat{R}$ = potential scale reduction factor (Gelman & Rubin, 1992); p_{T1} = posterior predictive p-value for the mean; p_{T2} = posterior predictive p-value for the covariance.

Supplemental Materials

accompanying the manuscript *Does Sleep Benefit Source Memory? Investigating 12-hr Retention Intervals with a Multinomial Modeling Approach*

Experiment 1

Experiment 2

Table S1

Sample Characteristics for Experiment 1

Variable	Total		Wake		Sleep			
	M	*SD*	*M*	*SD*	*M*	*SD*	*t*(130)	*p*
Age, in years	26.77	4.48	27.11	4.38	26.45	4.58	0.85	.399
SSS, learning session	2.48	0.96	2.55	0.85	2.40	1.06	0.90	.369
SSS, testing session	2.55	1.13	2.62	0.84	2.48	1.35	0.70	.485
rMEQ, sum score	14.00	3.83	14.15	3.79	13.85	3.89	0.45	.651
Retention interval	11.99	0.14	11.97	0.13	12.01	0.15	1.44	.151
Sleep duration					7.37	1.19		
Sleep quality					2.16	0.91		

Note. SSS = Stanford Sleepiness Scale (Hoddes et al., 1973); rMEQ = German version of the reduced Morningness-Eveningness Questionnaire (Randler, 2013). The SSS captures sleepiness on a 7-point scale with higher values indicating greater sleepiness. The rMEQ assesses chronotype with sum scores that are based on five items (Cronbach's α = .74). Sum scores can range from 4 to 25 with higher values indicating stronger morning preferences. The retention interval and self-reported sleep duration (*Mdn* = 7.50 hr, range: 3.00–9.33) in the experimental night are shown in hours. Sleep quality for the experimental night was assessed via self-report on a 5-point scale with higher values indicating better sleep quality. The wake condition contains 65 participants (41 female) of which 40 participants were recruited via Prolific and received payment whereas the sleep condition contains 67 participants (43 female) of which 46 participants were recruited via Prolific. The number of payed participants in the wake and sleep condition does not differ significantly, *t*(130) = 0.85, *p* = .395. In total, two participants (n_{wake} = 1, n_{sleep} = 1) hold an intermediate school certificate and 130 participants (n_{wake} = 64, n_{sleep} = 66) hold a qualification for college/university entrance. Whereas 78 participants are college/university students (n_{wake} = 38, n_{sleep} = 40), 54 participants (n_{wake} = 27, n_{sleep} = 27) pursue an occupation.

Table S2

Means and Standard Errors for the Memory Performance Measures in Experiment 1 by Type of Compensation

Dependent variable	No payment		Payment			
	Wake	Sleep	Wake	Sleep	*F*(1,128)	*p*
Hit rate	0.72 (0.02)	0.71 (0.02)	0.69 (0.01)	0.73 (0.01)	0.47	.496
False-alarm rate	0.05 (0.01)	0.07 (0.01)	0.12 (0.13)	0.07 (0.01)	4.63	.033
Sensitivity index *d′*	2.37 (0.08)	2.21 (0.08)	1.84 (0.09)	2.29 (0.08)	3.18	.077
Response bias *c*	0.49 (0.04)	0.44 (0.04)	0.36 (0.03)	0.43 (0.03)	0.66	.419
$\text{ACSIM}_{\text{Position}}$	0.72 (0.01)	0.75 (0.01)	0.67 (0.01)	0.78 (0.01)	2.32	.130

Note. Standard errors of the mean are presented in parentheses. Participants were either recruited via Prolifc (86; n_{wake} = 40, n_{sleep} = 46) or other channels (46; n_{wake} = 25, n_{sleep} = 21) and were randomly assigned to a wake versus sleep condition. Only participants recruited via Prolific were paid. ACSIM = average conditional source identification measure; *F* = *F* value for the interaction of compensation type and wake versus sleep.

Table S3

Sample Characteristics for Experiment 2

Variable	Total		Wake		Sleep			
	M	*SD*	*M*	*SD*	*M*	*SD*	*t*(132)	*p*
Age, in years	25.58	4.51	25.11	4.49	25.99	4.52	1.12	.265
SSS, learning session	2.74	1.27	2.68	1.13	2.79	1.38	0.52	.605
SSS, testing session	2.99	1.37	2.98	1.26	3.00	1.47	0.07	.946
rMEQ, sum score	13.08	4.04	12.95	3.62	13.19	4.39	0.35	.726
Retention interval	12.01	0.20	11.98	0.22	12.03	0.17	1.64	.104
Sleep duration					6.99	1.48		
Sleep quality					2.26	0.89		

Note. SSS = Stanford Sleepiness Scale (Hoddes et al., 1973); rMEQ = German version of the reduced Morningness-Eveningness Questionnaire (Randler, 2013). The SSS captures sleepiness on a 7-point scale with higher values indicating greater sleepiness. The rMEQ assesses chronotype with sum scores that are based on five items (Cronbach’s α = .77). Sum scores can range from 4 to 25 with higher values indicating stronger morning preferences. The retention interval and self-reported sleep duration (*Mdn* = 7.25 hr, range: 0.50–8.92) in the experimental night are shown in hours. Sleep quality for the experimental night was assessed via self-report on a 5-point scale with higher values indicating better sleep quality. The wake condition contains 62 participants (36 female) of which 43 participants were recruited via Prolific and received payment whereas the sleep condition contains 72 participants (48 female) of which 43 participants were recruited via Prolific. The number of payed participants in the wake and sleep condition does not differ significantly, *t*(132) = 1.16, *p* = .250. In total, the sample includes one pupil (sleep condition), two participants (n_{wake} = 1, n_{sleep} = 1) with no secondary school certificate, six participants (n_{wake} = 3, n_{sleep} = 3) with an intermediate school certificate and 125 participants (n_{wake} = 58, n_{sleep} = 67) with a qualification for college/university entrance. Whereas 70 participants are college/university students (n_{wake} = 33, n_{sleep} = 37), 64 participants (n_{wake} = 29, n_{sleep} = 35) pursue an occupation.

Table S4

Means and Standard Errors for the Memory Performance Measures in Experiment 2 by Type of Compensation

Dependent variable	No payment		Payment			
	Wake	Sleep	Wake	Sleep	*F*(1,130)	*p*
Hit rate	0.37 (0.02)	0.44 (0.02)	0.51 (0.02)	0.58 (0.02)	0.00	.944
False-alarm rate	0.13 (0.01)	0.09 (0.01)	0.14 (0.12)	0.15 (0.01)	1.42	.236
Sensitivity index *d′*	2.37 (0.08)	2.21 (0.08)	1.84 (0.09)	2.29 (0.08)	1.62	.205
Response bias *c*	0.49 (0.04)	0.44 (0.04)	0.36 (0.03)	0.43 (0.03)	0.84	.362
$\text{ACSIM}_{\text{Position}}$	0.59 (0.01)	0.70 (0.01)	0.61 (0.01)	0.64 (0.01)	2.66	.105
$\text{ACSIM}_{\text{Color}}$	0.58 (0.01)	0.61 (0.01)	0.53 (0.01)	0.53 (0.01)	0.30	.585

Note. Standard errors of the mean are presented in parentheses. Participants were either recruited via Prolifc (86; n_{wake} = 43, n_{sleep} = 43) or other channels (48; n_{wake} = 19, n_{sleep} = 29) and were randomly assigned to a wake versus sleep condition. Only participants recruited via Prolific were paid. ACSIM = average conditional source identification measure; *F* = *F* value for the interaction of compensation type and wake versus sleep.

References

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

Randler, C. (2013). German version of the reduced Morningness-Eveningness Questionnaire (rMEQ). *Biological Rhythm Research, 44*(5), 730–736. https://doi.org/10.1080/09291016.2012.739930

This dissertation was supported by the *Graduate School of Economic and Social Sciences (GESS)* at the University of Mannheim.