The current manuscript has been accepted for publication in *Psychological Bulletin*.

Citation:

Berres, S., & Erdfelder, E. (in press). The Sleep Benefit in Episodic Memory: An Integrative Review and a Meta-Analysis. *Psychological Bulletin*.

The supplemental materials accompanying this manuscript are provided here:

https://osf.io/h6fcx/?view_only=b639d876ce5147d89579498a83567ae0

© 2022, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/bul0000350

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

Author Note

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

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

Additional materials (i.e., codebook, data set, R script) are provided online on the Open Science Framework (OSF; https://osf.io/dr6hw/?view_only=c55b3737b6c642cfb170276088930a04). Based on a preliminary data set, a part of the findings was presented as a poster at the 2018 Conference of the German Psychological Society, Frankfurt am Main, Germany. Furthermore, parts of the results were presented in a talk at the 2021 joint annual meeting "Psychology and the Brain (PuG)" of the Division of Biological Psychology and Neuropsychology of the German Psychological Society and the German Society for Psychophysiology and its Application, Tübingen, Germany (online format). 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 Almut Baronius for coding the data as second coder and Lena Bizer for additional screening of titles and abstracts. Moreover, our thanks go

to Frank Renkewitz for his advice on the assessment of selective reporting. Finally, we thank all authors who provided us with all the information we needed.

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 C 207,

68159 Mannheim, Germany. Email: sberres@mail.uni-mannheim.de or erdfelder@psychologie.unimannheim.de

THE SLEEP BENEFIT IN EPISODIC MEMORY

3

Abstract

People recall more information after sleep than after an equally long period of wakefulness. This

sleep benefit in episodic memory has been documented in almost a century of research. However,

an integrative review of hypothesized underlying processes, a comprehensive quantification of the

benefit, and a systematic investigation of potential moderators has been missing so far. Here, we

address these issues by analyzing 823 effect sizes from 271 independent samples that were reported

in 177 articles published between 1967 and 2019. Using multilevel meta-regressions with robust

variance estimates, we found a moderate overall sleep benefit in episodic memory (g = 0.44).

Moderator analyses revealed four important findings: First, the sleep benefit is larger when stimuli

are studied multiple times instead of just once. Second, for word materials, the effect size depends

on the retrieval procedure: It is largest in free recall, followed by cued recall and recognition tasks.

Third, the sleep benefit is stronger in pre-post difference measures of retention than in delayed

memory tests. Fourth, sleep benefits are larger for natural sleep and nighttime naps than for

alternative sleep-study designs (e.g., SWS-deprived sleep, daytime naps). Although there was no

obvious evidence for selective reporting, it is a potential threat to the validity of the results. When

accounting for selective reporting bias, the overall effect of sleep on episodic memory is reduced but

still significant (g = 0.28). We argue that our results support an integrative, multi-causal theoretical

account of sleep-induced episodic memory benefits and provide guidance to increase their

replicability.

Keywords: episodic memory, sleep, consolidation, interference, meta-analysis

Public Significance Statement

This meta-analysis investigates effects of sleep versus wakefulness on episodic memory for previously learned information and its context (e.g., location, time). We found a moderate overall sleep benefit that increases (a) when items are studied repeatedly at encoding, (b) when using memory tests for verbal material that lack retrieval cues, (c) when employing pre-post difference measures of retention, and (d) when applying specific sleep study designs such as comparisons of nighttime sleep with daytime wakefulness. Overall, these results suggest that sleep strengthens both storage and retrieval of episodic memories.

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

Episodic memory is the memory for events, experiences or information and their respective context (e.g., location, time; Tulving, 2002). As such, this memory system comprises the ability to encode, store, and retrieve episodic information. Since the pioneering work of Jenkins and Dallenbach in 1924, there has been an ongoing interest in whether, why, and under which conditions sleep benefits episodic memory. Almost a century later, a considerable amount of research has accumulated showing that episodic memory indeed profits from sleep compared to wakefulness after learning. However, despite significant advances and important insights in the past decades, three important points have not yet been sufficiently clarified so far.

First, there is still no general agreement on the neurocognitive mechanisms that causally contribute to the sleep benefit in episodic memory. Although several theories on sleep effects have been proposed in psychology and neuroscience across the decades, the explanation of the sleep benefit is still a subject of a timely and lively debate (e.g., Antony & Schapiro, 2019; Yaroush et al., 1971; Yonelinas et al., 2019). As such, knowledge of the underlying processes provides valuable information about memory formation and has far-reaching implications for aging and mental health (e.g., Astill et al., 2012; Feld & Diekelmann, 2020; Goerke et al., 2017; Goldstein & Walker, 2014; 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). Therefore, it is crucial to better understand which mechanisms and processes contribute to the sleep benefit.

Second, although the sleep benefit in episodic memory can be considered a well-established finding, the typical size of the benefit across studies is still unclear. Knowledge of the effect size is indispensable when assessing the statistical power of published research and when planning future studies on the sleep benefit that are sufficiently powered. Moreover, the practical importance of the sleep benefit, for example, for aging and mental health research, crucially hinges on the typical strength of this effect. For all these reasons, it is essential to learn more precisely how much sleep actually improves episodic memory compared to wakefulness.

6

Third, for the primary studies that have accumulated so far, it is still unclear which variables moderate the sleep benefit significantly—if any. Researchers have used a wide variety of populations, materials, study designs, and dependent measures to investigate the sleep benefit. For example, the effect has been investigated in different age groups (e.g., children, older adults), for several sleep study designs (e.g., naps, total sleep deprivation), and for a variety of retrieval procedures and measures (e.g., cued recall, recognition; for a review, see Diekelmann et al., 2009). Apart from variability due to sampling and measurement error, is the assumption of a homogeneous sleep benefit across all these studies tenable? Or is there more evidence for systematic variability caused by some moderator variables? Knowledge about variables that moderate the sleep benefit helps researchers not only to better understand the processes underlying the sleep benefit and to test corresponding theories, but also to explain unexpected or even conflicting results and to design informative replication studies. Consequently, it is of prime importance to learn about variables that moderate the sleep benefit in episodic memory.

In what follows, we will address these three points step by step. We begin by reviewing prominent theories on the sleep benefit originating from psychology and neuroscience and classify them into two broad classes: consolidation and interference theories. Whereas consolidation theories focus on active processes that improve memory storage, most interference theories highlight the importance of memory retrieval. Both accounts are then used to derive hypotheses on potential moderators of the sleep benefit. Next, we quantify the sleep benefit in episodic memory with an extensive up-to-date meta-analysis of all relevant studies we could identify in the literature. This meta-analysis is based on 271 independent samples from more than 170 scholarly articles published in the past 52 years. Subsequently, using multilevel meta-regressions, we conduct and report separate moderator analyses. Specifically, we test hypotheses for 12 potential moderator variables to evaluate the two major theoretical accounts of the sleep benefit (i.e., consolidation and interference theories). Furthermore, to assess the potential impact of methodological decisions and reporting standards in primary studies, we investigate whether features of the study design and the

way effect sizes are reported or calculated by us affect the results. We also perform several sensitivity analyses to check the robustness of our findings. In addition, we assess the possible impact of selective reporting (i.e., publication bias) and provide a bias-corrected estimate of the overall effect of sleep on episodic memory. Finally, we critically discuss how consolidation and interference theories can account for the results of the current meta-analysis. To foreshadow, on the basis of our findings we propose a theoretical framework that integrates both classes of theories.

This framework allows for a broader and more comprehensive understanding of the processes jointly contributing to the sleep benefit in episodic memory. Therefore, it is a promising frame of reference to guide future research, to stimulate methodological innovations, and to derive additional predictions.

In sum, the contributions of the current review and meta-analysis are both practical and theoretical. We present evidence on the practical significance of the sleep benefit and provide guidance to increase replicability of future research in this area. Beyond this practical contribution, we develop a framework by integrating prominent theories from different research areas. As such, this integrative framework provides a broader and more comprehensive explanation of the sleep benefit in episodic memory than more fine-graded theories did before. In the following section, we first briefly summarize core assumptions of consolidation and interference theories before we review evidence from neuroimaging and polysomnography research that pinpoints neurocognitive mechanisms underlying sleep effects on memory storage in more detail. We then use consolidation and interference theories to derive hypotheses on potential moderators of the sleep benefit in episodic memory. Subsequently, these hypotheses will be tested in a meta-analysis of behavioral data on the sleep benefit in episodic memory that covers the empirical literature of the past five decades and beyond.

Processes Underlying the Sleep Benefit in Episodic Memory

Theories about the processes underlying the sleep benefit in episodic memory have been strongly influenced by the seminal work of Müller and Pilzecker (1900). In a series of studies, these

authors investigated learning and forgetting of nonsense syllables. On the basis of their results, they proposed that sufficiently strong memory representations of learned information persist and become firmly established after some time (perseveration-consolidation hypothesis). Moreover, they also discovered that memory representations are weakened when introducing interfering information after learning (retroactive interference hypothesis). In the following years, both hypotheses fostered a surge of research and were developed further by integrating other theories and findings (e.g., Hebb, 1949; Underwood, 1957). With regard to the sleep benefit in episodic memory, modified versions of both hypotheses have served as explanations. In what follows, we use "consolidation theories" as an umbrella term for all theories loosely connected to the perseveration-consolidation hypothesis, and "interference theories" to subsume further developments and improvements of the retroactive interference hypothesis.

Consolidation Theories

Memory consolidation refers to a process that converts 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 basically two forms of memory consolidation: Synaptic consolidation and systems consolidation. Synaptic consolidation refers to synaptic changes in local neural networks that stabilize memory representations within a few hours after encoding. In contrast, systems consolidation takes more time to be completed (e.g., weeks or months). It relies on synaptic plasticity (i.e., changes in synaptic strength; Dudai, 2004), which in turn depends on synaptic consolidation. Moreover, systems consolidation is additionally characterized by synaptic changes across various brain areas. Thereby, memory representations are not only stabilized but also qualitatively transformed. This qualitative transformation produces more abstract, generalized memory representations by gradually integrating various overlapping experience-based synaptic changes in the neocortex.

¹ Results from animal studies showed that systems consolidation of spatial memory in rats takes at least two weeks (Barry et al., 2016), but can also be found within 48 hr after learning if an appropriate mental schema exists prior to stimulus encoding (Tse et al., 2007).

9

In essence, both synaptic and systems consolidation are based on synaptic changes. These changes encompass synaptic up- and downscaling that can lead to either strengthening or weakening of existing connections but also to formation of new connections or to complete elimination of existing ones (for a literature review of empirical findings, see Puentes-Mestril & Aton, 2017). Synaptic upscaling as such hinges on replay, which is defined as the reactivation of specific neuronal firing patterns (Klinzing et al., 2019; Lewis & Durrant, 2011; O'Neill et al., 2010; Pfeiffer, 2020; Wilson & McNaughton, 1994). Although replay during retention has been observed in both sleep and wakefulness, sleep replay seems to differ fundamentally from wake replay (Carr et al., 2011; Dudai, 2012; Pfeiffer, 2020). For instance, more forward replay has been found in sleep than in wakefulness, whereas reverse replay has been primarily detected during wake phases (e.g., Foster & Wilson, 2006; Lee & Wilson, 2002; Wikenheiser & Redish, 2013). Moreover, results from animal studies showed impaired long-term memory in rats after blocking replay during sleep but not during wakefulness (e.g., Ego-Stengel & Wilson, 2010; Girardeau et al., 2009; Jadhav et al., 2012). Taken together, these findings point to a special role of sleep in memory consolidation (Dudai, 2012; Lewis & Durrant, 2011; Pfeiffer, 2020).

Various theories explain the sleep benefit in episodic memory by consolidation principles. One of these theories is the sequential hypothesis. It states that memories consolidate 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 example 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). Here, we focus on yet another theoretical account, namely, the active systems consolidation hypothesis (Born & Wilhelm, 2012; Diekelmann & Born, 2010; Feld & Born, 2017; Inostroza & Born, 2013; Klinzing et al., 2019; Rasch & Born, 2013). This hypothesis stands out because it incorporates principles of various other consolidation hypotheses including the sequential hypothesis and the synaptic homeostasis hypothesis. Therefore, the active systems consolidation hypothesis has been

considered "the currently most integrative account of sleep-dependent memory consolidation" (Klinzing et al., 2019, p. 1598).

According to the active systems consolidation hypothesis, components of a memory representation formed during wakefulness (e.g., color, texture, odor of a flower) are distributed across various brain areas and bound together into one unique episodic memory representation by the hippocampus. This hippocampal memory representation is replayed during subsequent sleep, especially during SWS. Importantly, it is assumed that specific aspects of the encoding situation, such as encoding strength or salience, foster hippocampal replay during sleep (Antony & Paller, 2017; Klinzing et al., 2019; Stickgold & Walker, 2013). In parallel to the hippocampal replay, the separate components of the hippocampal memory representation are also reactivated in the various brain areas. These synaptic upscaling processes are embedded in global synaptic downscaling that diminishes neuronal firing rates during sleep to renormalize the strength of synaptic connections. Hence, the active systems consolidation hypothesis posits that memory consolidation during sleep involves synaptic consolidation as well as systems consolidation. Both forms of memory consolidation are based on local synaptic upscaling and global synaptic downscaling processes in the hippocampus and neocortical areas. However, the precise interplay and time course of synaptic and systems consolidation as well as potential moderators are 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). Nevertheless, it appears that sleep supports the binding of various components into a unique episodic memory representation for shorter retention intervals (e.g., 10 hr; Weber et al., 2014) and the integration into preexisting memory networks, resulting in more gist-like memory representations for longer retention intervals (e.g., 3 days; Deliens et al., 2013; Klinzing et al., 2019). In sum, consolidation theories in general, and the active systems consolidation hypothesis in particular, propose an active role of sleep such that sleep strengthens storage of episodic memories in the short run (e.g., within 10 hr) and transforms them into stable but more gist-like representations in the long run (e.g., within 3 days).

Interference Theories

Whereas consolidation theories propose an active role of sleep in episodic memory storage, some interference theories propose a passive role of sleep in episodic memory storage and others in episodic memory retrieval.

Passive effects on memory storage. The same hippocampal resources that are required for memory consolidation are thought to be also required for learning new information during wakefulness (Wixted, 2004, 2005). Because the hippocampal resources are limited, new learning during wakefulness can disturb the consolidation process for previously learned information that continues for some time after memory encoding. During sleep, however, new learning is virtually absent. Thus, more hippocampal resources are available for memory consolidation of information learned before sleep (Wixted, 2004, 2005). Consequently, sleep provides a beneficial environment for memory consolidation by reducing retroactive interference (Ellenbogen, Payne, & Stickgold, 2006; Mednick et al., 2011; Wixted, 2004, 2005). In this way, sleep passively supports successful memory storage. Aside from creating an ideal condition for memory consolidation, sleep is not ascribed any special contributions to consolidation processes (Ellenbogen, Payne, & Stickgold, 2006; Mednick et al., 2011). That is, the process by which sleep supports memory consolidation and thus episodic memory storage is the same as for any other state with reduced retroactive interference (e.g., quiet wake, alcohol, benzodiazepines; Ellenbogen, Payne, & Stickgold, 2006; Mednick et al., 2011; Wixted, 2004, 2005). In sum, this theoretical account proposes that hippocampal memory consolidation occurs whenever the opportunity is provided by reducing retroactive interference. To clearly distinguish this account from interference theories that propose a passive role of sleep in episodic memory retrieval, we adopt the term "opportunistic consolidation hypothesis" from Mednick et al. (2011). In what follows, we use this term to refer to passive beneficial effects of sleep on memory storage.

Passive effects on memory retrieval. Early interpretations of the sleep benefit in episodic memory maintained that sleep temporarily protects memories from interference (e.g., Ekstrand,

1967; Jenkins & Dallenbach, 1924). In their pioneering study, Jenkins and Dallenbach (1924) presented lists of nonsense syllables to two participants in a series of experiments. The experiments varied in the timing of learning (i.e., day or night), the retention interval (i.e., 1, 2, 4, or 8 hr), and whether participants slept or stayed awake during this time. In a free recall task, the authors found that more syllables were successfully recalled after sleep than after wakefulness. They interpreted this sleep benefit in terms of the interference theory of forgetting—a theory particularly prominent in the early 20th century (McGeoch, 1932; Underwood, 1957). According to this theory, forgetting is caused by interference from prior (proactive interference) or subsequent learning (retroactive interference). In contrast to the waking state, subsequent learning is virtually absent during sleep. Consequently, retroactive interference is reduced, which in turn aids successful memory retrieval after sleep.

The notion that retroactive interference is due to retrieval but not to storage problems is convincingly supported by Tulving and Psotka's (1971) findings, among others. In their experiment, participants learned up to six word lists. Each list contained words from six semantic categories.

After word-list learning, two free recall tasks were administered, followed by a final cued recall task. Whereas in free recall tasks participants lacked retrieval cues, semantic category labels were provided in the cued recall task as memory aids. The authors found a pronounced retroactive interference effect in free recall—a decline in remembered words from the first list that became more pronounced as the number of subsequently learned lists increased. Importantly, the same retroactive interference effect was not evident in the final cued recall test. Thus, words from the first list appear to be available in memory, but these words are accessible only in cued recall and not in free recall because the latter test format lacks appropriate retrieval cues. These and many related results strongly suggest that retroactive interference is caused by retrieval and not by storage problems. If storage problems were involved, one would expect retroactive interference to impair both free and cued recall performance.

The notion that retroactive interference is primarily due to retrieval problems is broadly accepted nowadays and in fact incorporated in several current memory models. For instance, based on the temporal-distinctiveness account, the Scale-Invariant Memory, Perception and LEarning model (SIMPLE, Brown et al., 2007) predicts that the access to stored memories can be disrupted by proactive and retroactive interference. Specifically, the smaller the temporal distance between to-be-remembered and interfering information, the more difficult it is to discriminate and subsequently retrieve the target information at test. However, compared to wakefulness after learning, sleep reduces the temporal proximity between to-be-remembered information and subsequent learning after sleep. This leads to a reduction of retroactive interference, which in turn makes it easier to distinguish and retrieve the to-be-remembered information (Brown et al., 2007; Ecker, Brown, & Lewandowsky, 2015; Ecker, Tay, & Brown, 2015).

Drawing on this explanation, the more recent contextual binding account extends the cause of interference from temporal proximity to content or context similarity in general (Yonelinas et al., 2019). Essentially, context is used as a generic term to describe any aspect of a specific learning situation (e.g., cognitive, emotional, spatial, or temporal information). According to the contextual binding account, retrieval of target information can be impaired by information learned before or after, provided its content or context features resemble those of the target information. Thus, just like temporal proximity, content or context similarity may foster retroactive interference. Because new learning is virtually absent during sleep, retroactive interference due to content or context similarity is reduced. This facilitates discrimination and retrieval of to-be-remembered information after sleep relative to wakefulness (Yonelinas et al., 2019).

To summarize the interference theories reviewed in this paragraph, irrespective of the source of interference (temporal proximity, content similarity or context similarity), it is assumed that sleep after learning improves memory retrieval by protecting memories from retroactive interference.

Because the majority of interference theorists argue that sleep boosts memory retrieval, we will use the term "interference theories" only for theories that propose a passive role of sleep in episodic

memory retrieval hereafter. Hence, whereas the opportunistic consolidation hypothesis proposes that sleep improves episodic memory *storage*, interference theories as defined here maintain that sleep improves episodic memory *retrieval*. Notwithstanding this important difference, both theoretical accounts agree on a passive role of sleep by temporarily protecting episodic memories from interference.

In what follows, we first review evidence from neuroimaging and polysomnography research to evaluate active versus passive sleep effects on memory storage, as proposed by consolidation theories and the opportunistic consolidation hypothesis, respectively. We then consider predictions of consolidation and interference theories with respect to memory storage and retrieval, in turn developing corresponding hypotheses for a meta-analysis of behavioral measures of the sleep benefit and its potential moderators.

Evidence From Neuroimaging and Polysomnography Research

A core assumption of consolidation theories is that memory representations are replayed during sleep, especially during SWS. Thus, sleep—and particularly SWS—should possess specific features that are important for memory consolidation. In contrast, the opportunistic consolidation hypothesis states that memory consolidation occurs in any state that reduces retroactive interference (e.g., quiet wake, sleep). Thus, neither sleep, SWS nor quiet wake should possess any specific features that facilitate memory consolidation beyond reducing retroactive interference.

Neuroimaging techniques such as functional magnetic resonance imaging (fMRI) have greatly improved our understanding of whether the contribution of sleep to memory consolidation is unique or comparable to quiet wake. In fMRI studies, memory consolidation (i.e., systems consolidation) is indicated by changes in hippocampal-neocortical interactions between a recording before and after sleep or quiet wake. In fact, these studies show clear evidence for memory consolidation during sleep (e.g., Himmer et al., 2019; Sekeres et al., 2018; Sneve et al., 2015; Takashima et al., 2006; Takashima et al., 2009; Tompary et al., 2015). However, there is also some evidence for memory consolidation during quiet wake or minor cognitive activity such as judging whether a number is odd

or even (e.g., Staresina et al., 2013; Tambini et al., 2010; Wang et al., 2012). Thus, neuroimaging techniques provide evidence for both consolidation theories and the opportunistic consolidation hypothesis.

To investigate whether SWS possesses specific features that are important for memory consolidation as predicted by consolidation theories, the brain's electric activity during SWS was studied using electroencephalography in polysomnographic studies. Polysomnography (PSG) combines electroencephalography (EEG), electrooculography (EOG) and electromyography (EMG) to determine sleep stages (e.g., SWS, REM sleep). In fact, these studies identified the coupling of three specific SWS features (i.e., slow waves, sleep spindles, and sharp-wave ripples) as a key mechanism for memory consolidation during sleep (e.g., Helfrich et al., 2019; Staresina et al., 2015). To elaborate, slow waves are neocortical oscillations with high amplitudes and a frequency of 0.50 to 4.00 Hz. They comprise slow oscillations (0.50 to 1.00 Hz) as well as delta waves (1.00 to 4.00 Hz) and are involved in the synchronization of brain activity which is thought to facilitate memory consolidation (Barham et al., 2016; Klinzing et al., 2019; MacDonald & Cote, 2021; Muehlroth et al., 2020; Steriade, 2006). Sleep spindles are burst-like sequences of activation with a frequency of 10.00 to 16.00 Hz and a duration of 0.50 to 2 s. They originate from the thalamus and spread to the neocortex and hippocampus. Because sleep spindles appear to be involved in the redistribution and integration of memory representations, they are thought to promote memory consolidation (Fernandez & Lüthi, 2020; Klinzing et al., 2019; MacDonald & Cote, 2021; Muehlroth et al., 2020; Peyrache & Seibt, 2020; Steriade, 2006). Sharp-wave ripples are high frequency oscillations of 100 to 300 Hz (i.e., ripples) that follow a large deflection (i.e., sharp wave). They occur predominantly in the hippocampus during SWS but also in other brain areas and during quiet wake. It is assumed that sharp-wave ripples are related to synaptic down-selection and reactivation of neuronal firing patterns (Atherton et al., 2015; Girardeau & Zugaro, 2011; Klinzing et al., 2019; MacDonald & Cote, 2021; Muehlroth et al., 2020; O'Neill et al., 2010; Tononi & Cirelli, 2020). Further support for the predictions of consolidation theories comes from recent EEG studies that directly compared sleep

and quiet wake. Specifically, Zhang et al. (2018) showed that ripple-triggered replay occurs specifically during sleep. In addition, Nissen et al. (2021) found a positive correlation between slow waves and performance in a visual texture discrimination task only for a larger amount of non-rapid eye movement (NREM) sleep. Taken together, there is convincing evidence from polysomnography that SWS-specific features are important for memory consolidation as predicted by consolidation theories although some related neuronal processes also occur during quiet wake.

The prediction of the opportunistic consolidation hypothesis that memory consolidation occurs in any state that reduces retroactive interference receives some support from behavioral data.

Specifically, it has been shown that memory benefits are not specific to sleep or SWS but also occur during quiet wake (e.g., Humiston et al., 2019; Wamsley, 2019; Wang et al., 2021). However, there is also behavioral evidence from targeted memory reactivation (TMR) studies pointing to a special role of sleep in memory consolidation. As such, TMR studies experimentally investigate memory reactivation by presenting subtle auditory, olfactory, or tactile reminder cues during sleep or wakefulness. During prior learning, these cues were presented jointly with the learning material that is tested following a retention interval filled with either sleep or wakefulness. Cue presentation between learning and testing should reactivate associated memories and lead to better memory performance at test. In support of consolidation theories, a recent meta-analysis showed that TMR was only effective in stage 2 NREM sleep and SWS (Hu et al., 2020). This is in line with behavioral evidence from split-night studies. For example, Plihal and Born (1997) found that sleep in the first half of the night—predominantly characterized by SWS—exhibits a substantial sleep benefit in a

² To inform about the sleep benefit relative to wake conditions that differ in their activity levels—(1) usual everyday activity, (2) controlled cognitive activity across the retention interval, (3) controlled cognitive activity during part of the retention interval, and (4) wakeful rest—we conducted an additional exploratory moderator analysis based on the meta-analytic data set we introduce in a subsequent section. Results are provided on the OSF (https://osf.io/dr6hw/?view_only=c55b3737b6c642cfb170276088930a04). Descriptively, the sleep benefit relative to wakeful rest is not smaller in this analysis than relative to more active wake states, suggesting that physiological processes associated with sleep rather than cognitive inactivity per se drive the consolidation benefit. Note, however, that the number of effect sizes and the associated number of degrees of freedom for wakeful rest is too small for solid conclusions (k = 18, m = 4, df = 2.98).

paired-associate learning task whereas sleep in the second half of the night—predominantly characterized by REM sleep—does not.

In sum, there is compelling evidence from neuroimaging, polysomnography, and TMR studies that supports the claim of an active role of sleep in memory consolidation as proposed by consolidation theories. The results also show that an additional passive role of sleep in memory consolidation as predicted by the opportunistic consolidation hypothesis cannot be ruled out completely. However, more research is needed to check whether findings from sleep versus quiet wake comparisons can be generalized to other states of reduced retroactive interference (e.g., alcohol, benzodiazepines). Because the bulk of neurocognitive evidence unambiguously shows that sleep compared to wakefulness actively strengthens episodic memory storage, we focus on consolidation theories hereafter. Note, however, that this does not preclude an additional beneficial role of sleep in memory retrieval as suggested by most interference theories. In the following section, we therefore consider potential moderators of the sleep benefit in behavioral data and derive hypotheses from consolidation and interference theories that can be used subsequently to evaluate both theoretical accounts.

Potential Moderators

As outlined in Diekelmann et al.'s (2009) literature review, the sleep benefit in episodic memory appears to be affected by a number of moderator variables. So far, however, their impact has not yet been studied in detail and with up-to-date meta-analytical methods. Aggregating empirical findings across various studies allows unique insights into the potential impact of specific variables that can hardly be obtained from primary studies considered in isolation. For example, it is extremely challenging if not impossible to investigate potential moderator effect patterns of different stimulus materials or sleep study designs in single experiments. It is thus vital to inspect potential effects of such moderators across studies with appropriate meta-analytical methods.

The current meta-analysis considers 18 potential moderator variables previously discussed in the literature, some of which are closely linked to theories of the sleep benefit whereas other are not.

Specifically, we derive hypotheses from both consolidation and interference theories for a total of 12 variables. For seven of the 12 variables, consolidation and interference theories differ in their predictions. These variables are called "focal variables" hereafter. Because consolidation and interference theories make similar predictions for the remaining five variables, we refer to them as "parity check variables". To investigate whether the sleep benefit in episodic memory is moderated by methodological aspects of the study design, we examine four additional variables. We call these "exploratory variables" because, in contrast to the focal and parity check variables, we did not have any clear-cut predictions concerning their effects. Moreover, with two final variables we check whether randomization in primary studies and the accuracy with which we were able to extract effect size estimates from primary studies can account for the strength of the observed sleep benefit. These final two moderators are referred to as "control variables".

The focal, parity check, and exploratory variables can be assigned to five topic areas: (a) population characteristics (age), (b) aspects of the encoding situation (emotionality, restudy, testing, feedback, encoding instruction, directed forgetting instruction), (c) aspects of memory assessment (retrieval procedure, 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], SWS, REM sleep). Thus, a wide range of potential moderators is investigated, although we restricted the meta-analysis to variables for which sufficient data is available to allow for a joint moderator analysis. In the following section, we discuss the moderator variables investigated in this meta-analysis in more detail.

Focal Variables

Emotionality. It is a common finding that emotional information is better remembered than neutral information. This can be explained by an advantage of emotional information in encoding and storage: At encoding, emotional information receives more attention and thus more elaboration than neutral information. In addition, emotional content is often accompanied by a release of stress hormones that increase amygdala activation. This in turn increases the activation in the

hippocampus and consequently boosts memory consolidation (Hamann, 2001). Correspondingly, consolidation theories propose enhanced hippocampal replay during sleep for emotional compared to neutral memory representations, resulting in a sleep-related storage advantage for emotional memories (Payne & Kensinger, 2010; Walker & van der Helm, 2009).

However, emotional information can differ on two continuous dimensions: arousal (activation – deactivation) and valence (pleasure – displeasure; Russell, 1980, 2003, 2009). In fact, the two dimensions are often confounded so that information with negative valence are more arousing. More importantly, although the amygdala seems to be involved in both negative and positive information processing, negative information appears to elicit higher amygdala activation than positive information (Hamann et al., 2002; Lane et al., 1997; but see Sergerie et al., 2008). As sketched above, the activation of the amygdala can enhance hippocampus activity, and thus boost memory consolidation during sleep, especially for negative compared to positive memories.

Therefore, according to consolidation theories, the sleep benefit in episodic memory storage should be largest for negative, followed by positive, and lastly neutral memories. By contrast, interference theories do not make any strong predictions of whether and how emotionality moderates the sleep benefit in episodic memory retrieval.

The effect of emotionality on the sleep benefit has been investigated in two recent meta-analyses: One meta-analysis focused on recognition tasks and found no significant difference in the sleep benefit for emotional ($d_{SMC}^3 = 0.47$, 95% CI [0.29, 0.65], p < .001, k = 22) and neutral memories ($d_{SMC} = 0.47$, 95% CI [0.30, 0.63], p < .001, k = 22; Schäfer et al., 2020). Similarly, another meta-analysis, using combined effect sizes from free recall and recognition tasks, found comparable differences in sleep and wake conditions for negative and neutral memories (sleep: d = 0.52, 95% CI [0.28, 0.76], k = 20; wake: d = 0.58, 95% CI [0.40, 0.75], k = 20), positive and neutral memories (sleep: d = 0.22, 95% CI [-0.18, 0.62], k = 7; wake: d = 0.12, 95% CI [-0.13, 0.38], k = 7), and for emotional and neutral memories (sleep: d = 1.35, 95% CI [0.82, 1.89], k = 4; wake: d = 1.33, 95% CI [0.99, 1.67], k = 1.35, 95% CI [0.99, 1.67], k = 1.35, 95% CI [0.99, 1.67], k = 1.35, 95% CI [0.82, 1.89], k = 4; wake: k = 1.33, 95% CI [0.99, 1.67], k = 1.35, 95% CI [0.99, 1.67],

³ Effect size for pretest-posttest-control designs based on the difference between the standardized mean change (SMC) in a treatment and control group as proposed by Morris (2008).

4; Lipinska et al., 2019). These findings are at odds with the prediction of consolidation theories that the sleep benefit in episodic memory is largest for negative, followed by positive, and lastly neutral memories. However, the results of the two meta-analyses are not fully conclusive, because the impact of negative, positive, and neutral memories on the sleep benefit was not disentangled while controlling for other potential moderators. Here, we extend previous findings by filling this gap.

Restudy. Information can be learned in various ways. For example, after initial learning, the information can be studied again (with restudy) or not (no restudy). As such, restudy just like other aspects of the encoding situation (e.g., testing, feedback) impact the item's memory strength at encoding.4 According to the bifurcation model (Halamish & Bjork, 2011; Kornell et al., 2011), initial learning increases memory strength for all items equally, given that item difficulty is comparable. Importantly, restudy further increases memory strength for all items. 5 Because memory strength is a latent variable that is often assumed to be continuous and approximately normally distributed across all studied items, restudy can be understood as a shift of the memory strength distribution towards a higher level (Halamish & Bjork, 2011; Kornell et al., 2011). By implication, an item's memory strength cannot be measured directly by memory tests (e.g., free recall, cued recall, recognition). Rather, these memory tests reflect whether the item's memory strength is sufficiently high to surpass the test-specific retrieval threshold, resulting in a successful retrieval (Halamish & Bjork, 2011; Kornell et al., 2011; Rickard & Pan, 2018). Importantly, the retrieval threshold depends on the difficulty of the memory test, defined "in terms of its retention criterion, that is, the minimum memory strength at the end of the practice phase that is required for an item to be retained on the final test" (Halamish & Bjork, 2011, p. 804). On the basis of this definition, memory test difficulty increases not only with (a) weaker retrieval cues in memory tests (e.g., in free recall compared to cued recall; Halamish & Bjork, 2011), but also with (b) longer retention intervals (Kornell et al., 2011)

⁴ To inform about the combined effect of restudy, testing, feedback, and encoding instruction on the sleep benefit we provide the results of an additional exploratory moderator analysis on the OSF (https://osf.io/dr6hw/?view_only=c55b3737b6c642cfb170276088930a04).

⁵ Correspondingly, the dual memory model assumes that a study-memory is encoded during initial learning and strengthened by subsequent restudy (Rickard & Pan, 2018).

and (c) retroactive interference between original learning and delayed testing as a consequence of interpolated learning (Halamish & Bjork, 2011). Regarding restudy, this implies that after restudying compared to initial learning, more items exceed the retrieval threshold and can thus be successfully retrieved in a later memory test.

There is an ongoing debate about whether memory consolidation during sleep improves storage of weaker memories more than storage of stronger memories or vice versa (e.g., Born & Wilhelm, 2012; Diekelmann & Born, 2010; Diekelmann et al., 2009; Inostroza & Born, 2013; Klinzing et al., 2019; Stickgold, 2009). Whereas some empirical findings indicate that weaker memories profit more from memory consolidation during sleep (e.g., Denis et al., 2020; Drosopoulos et al., 2007), others show that stronger memories benefit more (e.g., Tucker & Fishbein, 2008; Schoch et al., 2017). These apparently conflicting findings can be reconciled by assuming that the sleep benefit follows an inverted U-shaped function of memory strength at encoding (Stickgold, 2009). According to this account, the sleep benefit in episodic memory storage should increase with memory strength until a medium level has been reached. In fact, growing empirical evidence has shown 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). More specifically, floor effects might counteract memory consolidation benefits during sleep when encoding strength is very low (Petzka et al., 2021). In contrast, when encoding strength increases beyond a medium level up to very high levels, then sleep benefits should decrease again. Hence, when memory strength is manipulated within this upper range only, sleep-induced storage improvements will be more pronounced for weaker compared to stronger encoding conditions. Such a locally detrimental effect of memory strength is plausible from the perspective of consolidation theory because very strong encoded memories already possess sufficient memory strength by default and therefore benefit less from further strengthening through memory consolidation during sleep (Denis et al., 2020; Denis et al.,

⁶ Note that this conception of a retrieval threshold differs fundamentally from the definition of a response criterion in Signal Detection Theory (SDT). Whereas SDT refers to *memory strength at test*, the bifurcation model refers to *memory strength at encoding* required for successful retrieval in a specific subsequent memory test.

2021; Petzka et al., 2021). Indeed, this assumption has been supported by EEG findings showing that higher spindle density during NREM sleep is associated with weaker encoded memories (Denis et al., 2021; Schmidt et al., 2006). In line with this, other studies have shown that TMR during sleep yielded better memory performance in a delayed test for weaker compared to stronger encoded memories (Cairney et al., 2016; Creery et al., 2015).

Despite the growing support for an inverted U-shaped curve of the sleep benefit in episodic memory storage, there is also recent behavioral evidence supporting the view 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 weaker and stronger memories in two conditions: (a) a relatively easy standard-test condition that included a delayed memory test after the retention interval, and (b) a difficult-test condition that additionally introduced retroactive interference before the delayed memory test. A sleep benefit was observed for weaker but not for stronger memories in the standard-test condition. In the difficult-test condition, however, a sleep benefit occurred for both weaker and stronger memories. A possible interpretation is that the sleep benefit in episodic memory storage occurs irrespective of memory strength unless it is concealed by ceiling effects for strong memories in the relatively easy standardtest condition (Petzka et al., 2021). In summary, the empirical evidence concerning encoding strength effects on memory consolidation during sleep is quite mixed and more rigorous testing of the assumed inverted U-shaped curve is needed. Future research should therefore directly assess the sleep benefit as a function of encoding strength by systematically manipulating strength in small steps across a broad range, ideally using the same encoding strength manipulation. This notwithstanding, the theoretical predictions of consolidation theories are clear: Up to a medium level of encoding strength, sleep benefits in memory storage should increase with memory strength, and these benefits should decrease with memory strength once a medium strength level is exceeded.

What does this mean with respect to predicted effects of the potential moderator variable "restudy"? Restudying items increases memory strength of all studied items alike. Most studies included in the current meta-analysis probably ensure relatively high memory strength at encoding to find a sleep benefit for both restudied and not restudied items. If so, consolidation theories would predict smaller sleep benefits for restudied compared to not restudied items because encoding conditions typically vary on the upper limb of the inverted U-shaped function. Note, however, that this remains speculation because the actual level of memory strength at encoding across all studies included in the meta-analysis is unknown. Thus, on the basis of the available information, clear-cut predictions from consolidation theories on whether and how restudy moderates the sleep benefit in episodic memory storage cannot be derived.

Whereas consolidation theories predict sleep-induced increases in memory strength due to memory consolidation, interference theories predict lower retrieval thresholds as a consequence of sleep-reduced retroactive interference. As noted before, whether an item is successfully retrieved in a memory test depends not only on its memory strength but also on the retrieval threshold and thus the difficulty of the memory test (Halamish & Bjork, 2011; Kornell et al., 2011; Rickard & Pan, 2018). There is ample empirical evidence showing that retroactive interference hinders memory retrieval (e.g., Ecker, Brown, & Lewandowsky, 2015; Ecker, Tay, & Brown, 2015; Halamish & Bjork, 2011; Petzka et al., 2021). Thus, reduced retroactive interference during sleep as predicted by interference theories should facilitate memory retrieval by lowering the retrieval threshold required for successful recall in a delayed test. Consequently, the retrieval threshold in the sleep condition should be lower than in the wake condition. Because of this threshold shift, sleep compared to wakefulness should boost memory retrieval. However, the observed sleep benefit in episodic memory not only depends on the lowering of the retrieval threshold, but also on the relative memory strength in weak- and strong encoding conditions. Whereas testing is assumed to increase memory strength of successfully tested items to a greater extend, restudy and feedback increase

memory strength to a lesser extend (e.g., Halamish & Bjork, 2011; Pastötter & Bäum, 2016; Kornell et al., 2011).

Based on these considerations, how should the potential moderator "restudy" affect the sleep benefit if interference theories hold? To reiterate, restudy compared to no restudy should increase memory strength for all studied items alike. However, because restudy increases memory strength only to a lesser extend (e.g., Halamish & Bjork, 2011; Pastötter & Bäum, 2016; Kornell et al., 2011), the memory strength distributions for restudy and no-restudy conditions are relatively close to each other. In addition, sleep compared to wakefulness reduces retroactive interference and thus lowers the retrieval threshold. Consequently, more restudied than not restudied items surpass the retrieval threshold in the sleep than in the wake condition (for an illustration, see Figure 1A). The sleep benefit in episodic memory retrieval should thus be larger for encoding contexts characterized by restudy than for those with a single learning trial only.

Testing. Encoding procedures differ not only in whether or not restudy options are provided, but also in whether retrieval of the learned information is practiced during the encoding phase (with testing) or not (no testing). The bifurcation model predicts that items successfully retrieved during retrieval practice have a greater memory strength than non-retrieved items (Halamish & Bjork, 2011; Kornell et al., 2011). This is because retrieval per se serves as a learning event (e.g., Halamish & Bjork, 2011; Kornell et al., 2011; Rowland, 2014). Specifically, during retrieval practice some items are successfully retrieved and gain memory strength, while other items are not retrieved and remain at the same level of memory strength as not tested items. Thus, for retrieval practice, the memory strength distribution is bifurcated after the learning phase (Halamish & Bjork, 2011; Kornell et al., 2011; Rickard & Pan, 2018). Note also that retrieval practice increases memory strength of successfully retrieved items to a larger extent than restudy or feedback (e.g., Halamish & Bjork, 2011; Pastötter & Bäuml, 2016; Kornell et al., 2011). Overall, in a later memory test, more items

⁷ Note that this is also predicted by the dual memory model (Rickard & Pan, 2018).

should exceed the retrieval threshold and thus be successfully retrieved when retrieval practice takes place during the learning phase than when it does not (see Figure 1B for an illustration).

For the potential moderator "testing", consolidation theories predict further sleep-induced strengthening of memory representations of tested and not tested items depending on memory strength at encoding. This resembles the prediction of consolidation theories for restudied and not restudied items. Because the actual level of memory strength at encoding varies across studies included in the current meta-analysis in unknown ways, we again cannot derive a clear-cut prediction from consolidation theories of whether and how testing moderates the sleep benefit in episodic memory storage.

Interference theories, in contrast, predict no substantial difference in the sleep benefit between testing and no testing conditions. To see this, recall that the memory strength distribution in the testing condition is bifurcated. That is, successfully retrieved items in the testing condition gain more memory strength than not retrieved items. Whereas the memory strength distribution in the testing condition is bifurcated, the approximately normal memory strength distribution in the no-testing condition remains unchanged. Thus, the memory strength distribution for not retrieved items in the testing condition and the memory strength distribution in the no-testing condition overlap (see Figure 1B). By contrast, the memory strength distribution for successfully retrieved items in the testing condition is shifted to a higher level of memory strength. Consequently, assuming the same retrieval thresholds as before, the proportion of items exceeding the lower sleep threshold compared to the higher wake threshold is quite similar in testing and no-testing conditions (see Figure 1B). In sum, interference theories predict that sleep compared to wakefulness boosts memory retrieval for testing and no-testing conditions to about the same extent.

Feedback. Encoding procedures also vary in whether feedback is provided after testing or not (with vs. no feedback, respectively). As such, feedback is generally considered as an additional learning opportunity (e.g., Butler et al., 2008; Roediger & Butler, 2011). By providing correct answers and reducing uncertainties, feedback is not expected to affect memory strength of successfully

tested items, but to increase memory strength of tested but not yet successfully retrieved items. Thus, feedback in comparison to no feedback after testing eliminates or reduces the bifurcation of the memory strength distribution for tested items (Kornell et al., 2011; Mundt et al., 2020; Pashler et al., 2005; Pastötter & Bäuml, 2016). In other words, if feedback is provided after testing, the memory strength distribution for not successfully retrieved items in the testing condition shifts to a higher level of memory strength which is close to the memory strength of successfully tested items. However, if no feedback is given, the memory strength distribution for not successfully retrieved items does not shift, resulting in the persistence of the bifurcation in the memory strength distribution due to testing (for an illustration, see Figure 1C).

Like for restudy and testing, consolidation theories predict further strengthening of memory representations in the sleep compared to the wake condition depending on memory strength at encoding for items with and without feedback. However, for the reason described in the paragraphs on restudy and testing, clear-cut predictions on whether and how feedback moderates the sleep benefit in episodic memory storage cannot be derived from consolidation theories.

Interference theories, however, predict a larger sleep benefit in feedback than in no-feedback conditions. As sketched above, feedback compared to no feedback increases memory strength of items not successfully retrieved in the testing condition, thereby eliminating or reducing the bifurcation in the memory strength distribution of tested items (Kornell et al., 2011; Mundt et al., 2020; Pashler et al., 2005; Pastötter & Bäuml, 2016). According to interference theories, sleep reduces retroactive interference which results in a lower retrieval threshold after sleep than after wakefulness. Consequently, more previously unretrieved items surpass the lower retrieval threshold after sleep if feedback compared to no feedback is provided in the encoding phase (see Figure 1C). Thus, sleep compared to wakefulness should improve memory performance, especially if feedback was provided after testing.

Encoding Instruction. Two types of encoding are typically distinguished: intentional learning versus incidental learning. In intentional learning, participants are explicitly instructed beforehand to

learn items for a later memory test. This memory test can follow right after learning or in a delayed memory test after a retention interval. In incidental learning, by contrast, participants are instructed to perform a task related to the items (e.g., rating the stimuli for likability) without mentioning the upcoming test that will follow either immediately or delayed as a surprise memory test. Hence, items are learned deliberately in intentional learning whereas they are learned incidentally without a specific intention to do so in incidental learning. Because encoding effort is typically higher during intentional learning compared to incidental learning, the former typically results in stronger memories than the latter (Noldy et al., 1990). By implication, encoding instructions may moderate the sleep benefit due to their potential impact on memory strength.

According to consolidation theories, intentionally and incidentally learned memories are further strengthened by memory consolidation depending on their encoding strength. Again, because the actual level of memory strength at encoding across studies included in the current meta-analysis is unknown, predictions on whether and how encoding instructions moderate the sleep benefit cannot be derived unambiguously.

Interference theories, however, clearly predict a larger sleep benefit in intentional than in incidental learning. Resembling restudy conditions, intentional compared to incidental learning should increase memory strength for all studied items alike. That is, the memory strength distribution following intentional learning should shift to a higher level, but still be relatively close to the corresponding distribution following incidental learning. Because sleep compared to wakefulness lowers the retrieval threshold by reducing retroactive interference, more intentionally learned items will surpass the retrieval threshold after sleep than after wakefulness, resulting in a larger sleep benefit for intentional than for incidental learning (see Figure 1A).

Directed Forgetting Instruction. Everyday life provides huge amounts of information that are not processed in equal depth due to limited cognitive resources. Indeed, favorable memory performance depends on remembering important or relevant information on the one hand and on forgetting unwanted or irrelevant information on the other hand (e.g., Han et al., 2020). To investigate

memory differences between information that is intended to be remembered versus forgotten, researchers use various paradigms (e.g., think/no-think, item-method directed forgetting, list-method directed forgetting; e.g., Bäuml et al., 2010). Across these paradigms, memory performance in a later memory test for items subsequently instructed to be remembered is typically much better than memory performance for items subsequently instructed to be forgotten. This finding is also known as the directed forgetting effect. However, there is an ongoing debate about the underlying cognitive processes, which likely differ between paradigms (e.g., Abel & Bäuml, 2019; Bäuml et al., 2010; Rummel et al., 2016).

According to consolidation theories, the intention to remember a learned information in the future should enhance hippocampal replay during sleep (Antony & Paller, 2017; Stickgold & Walker, 2013). Specifically, it has been proposed that during or shortly after encoding, important or relevant information is tagged for persistent synaptic strengthening, also known as long-term potentiation (Barco et al., 2008; Frey & Morris, 1998; Morris, 2006; Redondo & Morris, 2011). However, the mechanisms by which synaptic tags impacts hippocampal replay during sleep have not yet been specified clearly. Nevertheless, it has been hypothesized that synaptic tags lead to increased memory consolidation. Hence, memory storage of important or relevant information should benefit more than storage of unwanted or irrelevant information (Antony & Paller, 2017; Stickgold & Walker, 2013), leading to a larger sleep benefit for to-be-remembered than for to-be-forgotten information. Interference theories, in contrast, do not make any strong predictions of whether and how directed forgetting instructions moderate the sleep benefit in episodic memory retrieval.

Retrieval Procedure. There are multiple methods to assess episodic memory, for example free recall, cued recall, and recognition tasks. These retrieval procedures differ in how much they rely on memory retrieval. Whereas in free recall tasks no retrieval cues are provided, typically one cue is presented in cued recall and even more cues in recognition tasks. Thus, free recall tasks impose a much higher burden on memory retrieval than cued recall and recognition tasks.

According to consolidation theories, sleep compared to wakefulness boosts memory storage via memory consolidation. This should be reflected in measures of free recall, cued recall and recognition to about the same extent. In general, therefore, consolidation theories predict no substantial difference in the sleep benefit between these retrieval procedures.

Diekelmann et al. (2009), however, pointed to a possible side effect of memory consolidation that might explain a larger sleep benefit in recall (i.e., free and cued recall tasks combined) than in recognition tasks. The authors argued that during systems consolidation, memory representations of to-be-recalled information become connected to many cues as a consequence of integration into preexisting memory networks. In turn, these cues should facilitate access and thus memory retrieval. Because accessibility is more of a problem in recall compared to recognition, the observed sleep benefit should be larger in recall than in recognition tasks (Diekelmann et al., 2009). Although this hypothesis is plausible, it is challenged by empirical evidence showing that behavioral effects of systems consolidation take time to manifest. Although neuronal changes associated with systems consolidation already occur during learning (e.g., Brodt et al., 2018; Himmer et al., 2019; Kitamura et al., 2017), 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). Even though we cannot completely rule out a side effect of systems consolidation on memory retrieval, it seems safe to assume that such an effect is negligible in the current meta-analysis. Because the retention intervals are relatively short for the majority of studies included in the meta-analysis (Mdn = 12 hr), memory consolidation during sleep should primarily support memory storage, not memory retrieval. Given these conditions, consolidation theories would predict no substantial differences in the sleep benefit between free recall, cued recall, and recognition tasks.

In contrast, interference theories would predict a particularly pronounced sleep benefit for tasks that heavily rely on appropriate retrieval processes. Specifically, according to interference theories, sleep compared to wakefulness improves memory retrieval in a later memory test by reducing retroactive interference. Therefore, the sleep benefit should become more apparent the more

difficult the memory test is with respect to retrieval. Because free recall relies more on memory retrieval than cued recall, and cued recall more than recognition, the sleep benefit should be largest in free recall, followed by cued recall, and lastly recognition tasks (cf. Dyne et al., 1990; McKinney, 1935; Postman, 1952).

In summary, consolidation and interference theories differ in their predictions about whether the retrieval procedure moderates the sleep benefit in episodic memory. Whereas consolidation theories predict no substantial differences in the sleep benefit between different retrieval procedures at least for retention intervals up to 24 hr length, interference theories predict the largest sleep benefit for free recall, followed by cued recall, and lastly recognition tasks.

Parity Check Variables

Age. Most studies on the sleep benefit in episodic memory were conducted with young adults. Some studies, however, also investigated the sleep benefit in additional age groups, ranging from children and adolescents to older adults. Age may moderate the sleep benefit due to age related changes in sleep parameters. Specifically, these changes include a decline in total sleep time (TST) and SWS as well as an increase in sleep interruptions across the life span (Iglowstein et al., 2003; Ohayon et al., 2004; Skeldon et al., 2016). Major changes in sleep parameters typically occur around the ages of 10, 35, 55, and 75 years (Ohayon et al., 2004). To illustrate, TST declines around the age of 10 years before it levels out around the age of 35 years and stays relatively constant until 55 years. Furthermore, SWS declines approximately from the age of 10 years, whereas REM sleep increases around the age of 10 years and remains relatively constant before it drops around the age of 55 years. By contrast, wake after sleep onset appears to increase from the age of 35 years (for details see Ohayon et al., 2004).

Consolidation theories consider SWS important for the replay of hippocampal memory representations. Therefore, an age-related decline in SWS should reduce memory consolidation and thus limit the sleep benefit in episodic memory storage. By contrast, interference theories predict

⁸ Note that the decline in TST around the age of 10 years is probably due to environmental, not biological factors (e.g., early school start; Ohayon et al., 2004).

that shorter TST and more sleep interruptions increase retroactive interference, which in turn should reduce the sleep benefit in episodic memory retrieval. In sum, although consolidation and interference theories offer different explanations, both theoretical accounts would predict that the sleep benefit decreases with advancing age. Thus, the sleep benefit in episodic memory should be largest for children and adolescents, followed by adults, and lastly older adults.

This hypothesis is supported by a recent meta-analysis that investigated the impact of age on the sleep benefit in episodic memory by summarizing 20 effect sizes (k) from 15 articles (Gui et al., 2017). The authors observed better memory performance after sleep compared to wakefulness for young adults (ca. 18 – 35 years old; d = 1.14, 95% CI [0.59, 1.70], p < .001, k = 10) but not for older adults (ca. 60 – 85 years old; d = 0.25, 95% CI [-0.11, 0.65], p = .188, k = 10). In the present meta-analysis, we expand this work by including available data across the entire life span.

Definition of Dependent Variable. Researchers measure the sleep benefit in different ways.

Consider a typical experiment to investigate the sleep benefit in episodic memory. This experiment comprises two sessions that are separated by a retention interval filled with either sleep or wakefulness. In the first session, participants study the material and are perhaps immediately tested to assess their learning success. After the retention interval, participants perform a delayed test in the second session. Whereas some researchers use only the delayed test to compare memory performance after sleep and wakefulness, other researchers compute the difference between the immediate and delayed test. This change score is a direct measure of retention and controls for potential individual differences in learning and memory abilities. Therefore, it likely provides a more reliable measure of the sleep benefit compared to measures that are based on delayed tests only. On the basis of this reasoning, the sleep benefit for direct measures of retention (i.e., the difference immediate test - delayed test) should be larger than for measures of delayed test memory performance. This should apply regardless of the underlying processes of the sleep benefit proposed by consolidation and interference theories. By evaluating the cumulative evidence for this

⁹ Note that direct measures of retention are more precise measures of memory performance only if immediate and delayed test performance are positively correlated. This is usually the case in episodic memory tests.

prediction, we aim to advance our understanding of how methodological decisions concerning the research design may impact the sleep benefit in episodic memory.

TST. Studies vary in terms of the total amount of time participants in the sleep condition spent sleeping. Because memory consolidation and interference reduction could be affected by the time spent asleep, TST may moderate the sleep benefit. As such, sleep benefits have been observed for short (e.g., 6 min) as well as for longer sleep durations (e.g., 8 hr). Importantly, within the framework of consolidation theories, it has been speculated that longer sleep durations are more beneficial for memory consolidation than short sleep durations (Diekelmann & Born, 2010; Diekelmann et al., 2009). Thus, the sleep benefit in memory storage should increase with TST according to consolidation theories. ¹⁰ In addition, interference theories predict an increase of the sleep benefit with TST. Because new learning is virtually absent during sleep, retroactive interference should be reduced even further with longer TST. Hence, although consolidation and interference theories offer different explanations, both accounts predict that longer TST is associated with larger sleep benefits in episodic memory.

Two meta-analyses investigated the relationship between sleep duration and memory performance. One meta-analysis focused on 5 to 12 years old children and found no significant correlation between sleep duration and explicit memory (r = 0.07, 85% CI [0.00, 0.14], k = 11; Astill et al., 2012). A second meta-analysis compared short sleep duration (≤ 5 hr) and long sleep duration (≥ 9 hr) with normal sleep duration (7 - 8 hr) in older adults (i.e., ≥ 55 years old). This study showed that both short and long sleep duration were associated with poorer verbal memory (short sleep duration: OR = 1.33, 95% CI [1.19, 1.49], k = 23; long sleep duration: OR = 1.47, 95% CI [1.31, 1.66], k = 21; Lo et al., 2016). However, to our knowledge, there has not yet been a meta-analysis that explores TST as a potential moderator of the sleep benefit in episodic memory across age groups. In

¹⁰ It has also been speculated that longer sleep durations of more than 1 to 2 hr do not provide additional storage advantages (see Diekelmann et al., 2009). Therefore, we conducted a sensitivity analysis on the subgroup data set for TST to examine whether there is a quadratic instead of a linear relation between TST and the sleep benefit in episodic memory. There was no evidence for a quadratic trend. For details of this analysis see Table S55 in the supplemental materials.

the present meta-analysis, we aim to fill this gap by controlling for age, but also for other potential confounders such as the length of the retention interval and the sleep study design.

SWS. The time participants in the sleep condition spent in SWS may moderate the sleep benefit in episodic memory. According to consolidation theories, hippocampal memory representations are especially replayed during SWS (e.g., Diekelmann & Born, 2010; Inostroza & Born, 2013; Klinzing et al., 2019). Thus, more time spent in SWS should increase the sleep benefit in memory storage. Likewise, interference theories predict that the sleep benefit in memory retrieval increases with SWS duration. Because new learning is virtually absent during SWS, the reduction of retroactive interference should increase with more time spent in SWS. Taken together, both consolidation and interference theories predict that longer SWS duration increases the sleep benefit in episodic memory but offer different explanations. The current meta-analysis tests this prediction by controlling for the length of the retention interval and other potential confounders.

REM Sleep. The contribution of REM sleep to the sleep benefit in episodic memory remains controversial because it is difficult to dissociate REM sleep from other sleep stages and stress effects (Boyce et al., 2017; Diekelmann & Born, 2010; Diekelmann et al., 2009; Siegel, 2001). In fact, within the framework of consolidation theories, it has been speculated that sleep stages share specific neuronal processes. To elaborate, global synaptic downscaling should occur during both SWS and REM sleep, but synaptic upscaling only during SWS. Importantly, it is assumed that global synaptic downscaling is more pronounced during REM sleep compared to SWS (Klinzing et al., 2019).

Another assumption that specifically refers to episodic memory storage is that emotional memory representations are replayed during REM sleep theta activity and are thus protected from global downscaling (Hutchison & Rathore, 2015; Klinzing et al., 2019; Walker, 2009). This hypothesis is supported by studies indicating that REM sleep duration is associated with better memory for emotional compared to neutral information (e.g., Schäfer et al., 2020; for literature reviews see Tempesta et al., 2018; Walker & van der Helm, 2009). What is more, negative compared to positive memories seem to be associated with higher amygdala and thus hippocampus activation, which in

turn boosts memory consolidation during sleep. Thus, REM sleep may especially support storage of negative memories. Taken together, according to consolidation theories, the sleep benefit in memory storage should increase with REM sleep duration, especially for negative memories.

By contrast, interference theories predict that longer REM sleep increases the sleep benefit in memory retrieval, irrespective of the stimuli's emotionality. As such, REM sleep should temporarily protect episodic memories—irrespective of their emotionality—from retroactive interference.

In sum, although consolidation and interference theories offer different explanations, both theoretical accounts predict that more time spent in REM sleep increases the sleep benefit in episodic memory. Even though the question whether REM sleep contributions to the sleep benefit in memory storage differs between negative, positive, and neutral information is of considerable interest, we unfortunately cannot address it on the basis of the available data. For example, the data underlying our meta-analysis include only four studies (k = 8 effect sizes) that investigate positive memories and report the amount of REM sleep. We therefore focus on REM sleep duration as a potential moderator of the sleep benefit while controlling for the emotionality of the stimuli, among others.

Exploratory Variables

Stimulus Material. To investigate the sleep benefit in episodic memory, researchers use a wide range of stimulus materials. This includes (a) verbal material such as single words, word pairs, or prose passages; (b) nonverbal material such as pictures, colors, locations, sounds, or odors; and (c) various combinations of verbal and nonverbal materials. Despite the variety of stimulus materials used in previous research, stimulus material has not yet been investigated as a potential moderator of the sleep benefit to our knowledge. Therefore, the current meta-analysis investigates whether the sleep benefit in episodic memory generalizes across various stimulus materials.

Sleep Study Design. Differences between sleep and wakefulness have been investigated in many ways. In fact, sleep study designs range from comparisons of natural night sleep and daytime wakefulness, assessment of the effects of naps and total sleep deprivation to selective deprivation of

SWS or REM sleep. All sleep study designs differ with regard to various features such as the time at which participants sleep versus stay awake, the length of the retention interval, TST, and the amount of SWS as well as REM sleep. To test whether the sleep benefit in episodic memory differs between study designs, we investigate the sleep study design as a potential moderator.

Retention Interval. As sketched above, sleep study designs can differ in the length of the retention interval. Consider for example a nap study, where the retention interval between learning and a later memory test is 2 hr and a sleep deprivation study where the later memory test takes place after 2 days. In addition, the retention interval can also vary in length within a specific sleep study design. For example, in a nap study, the retention interval is 1 hr, whereas in another nap study, the retention interval covers 5 hr. As longer retention intervals are associated with more forgetting (Staugaard & Berntsen, 2019), the length of retention interval may moderate the sleep benefit in episodic memory.¹¹

Experimental Design. Researchers have compared sleep versus wakefulness between participants (i.e., participants either sleep or stay awake before their memory is tested) as well as within participants (i.e., all participants are tested after a period of sleep and after a period of wakefulness). From a statistical point of view, the within-subject design is more efficient than the between-subjects design, that is, it has higher power to detect differences between the sleep and wake condition. However, from a substantive point of view, the within-subject design may be problematic due to unwanted effects of learning, material, sensitization, consistency, or demand characteristics. To assess whether the sleep benefit in episodic memory differs between design types, we explore the experimental design as potential moderator.

Control Variables

We additionally incorporated two control variables into the moderator analysis. Specifically, we added a dummy variable "randomization" to indicate the certainty with which we were able to determine whether a study was truly randomized or not. Due to scheduling problems or missing

¹¹ Note that Staugaard and Berntsen (2019) investigated forgetting over time. They did not compare sleep versus wakefulness.

reports, it is not always clear whether random assignment of participants to sleep versus wake conditions took place. Similarly, we included a dummy variable "accuracy of effect size computation" to indicate the precision with which we were able to determine effect sizes in the primary studies. If we could not obtain the necessary information to calculate the effect size directly from the text or from tables, we used graphs and replacement statistics to estimate the effect size (e.g., the *p* value of a *t* test). Consequently, effect sizes might differ in the precision with which they reflect the actual observed effects if the corresponding effect sizes were not directly reported in the primary studies. Incorporating these two control variables into the moderator analysis helps to evaluate whether results are affected by uncertainties that occurred when coding primary studies.

The Current Meta-Analysis

Despite the wealth of qualitative and quantitative research syntheses on sleep and memory, the present meta-analysis is the first that explicitly focuses on episodic memory. Previous meta-analyses addressed, among others, prospective or motor memory (Leong et al., 2019; Pan & Rickard, 2015; Schmid et al., 2020), specific methods such as transcranial electrical stimulation or TMR (Barham et al., 2016; Hu et al., 2020), special populations (Astill et al., 2012; Gui et al., 2017; Lo et al., 2016; Wardle-Pinkston et al., 2019), experimental manipulations of sleep restrictions (Lowe et al., 2017) or emotionality (Lipinska et al., 2019; Schäfer et al., 2020), and research paradigms for item integration, rule extrapolation or false memories (Chatburn et al., 2014; Newbury & Monaghan, 2019). To provide a comprehensive analysis of sleep effects on episodic memory, we analyzed results published between 1967 and 2019 on healthy human individuals (a) to estimate the overall effect of sleep on episodic memory and (b) to systematically investigate potential moderator variables. Subsequently, we use these results to evaluate consolidation and interference theories of the sleep benefit in episodic memory.

Method

Additional materials for the current meta-analysis such as the codebook, data set, and R script, are available on the OSF (https://osf.io/dr6hw/?view_only=c55b3737b6c642cfb170276088930a04).

Tables S1 to S3 in the supplemental materials list the articles included in the meta-analysis along with their respective sample characteristics, moderator characteristics, and effect sizes.

Search Strategy

In January 2018, we systematically searched all relevant electronic databases (i.e., PsycInfo, PSYNDEX, ProQuest Dissertations and Theses Global, PubMed, Web of Science) for records published up to the search date that satisfied the search request (*sleep* OR *napping* OR *NREM sleep* OR *REM sleep* OR *polysomnography* OR *sleep deprivation* OR *sleep onset* OR *sleep wake cycle* OR *human biological rhythms* OR *chronotype*) AND (*episodic memory* OR *memory consolidation* OR *proactive inhibition* OR *retroactive inhibition*). To identify the appropriate search terms used in this request, we relied on the APA's *Thesaurus of Psychological Index Terms*. Subsequently, our systematic search was completed with an unsystematic search for additional relevant studies the authors became aware of until the end of 2018.

Study Selection

Study selection involved four steps summarized in the PRISMA flow diagram presented in Figure 2. In the first step 1,876 duplications were removed from the 5,278 identified records. During the second step, the first author screened the titles and abstracts of the remaining 3,402 records. In addition, a second coder screened the titles and abstracts of 851 records (25%) randomly selected from the 3,402 records, resulting in overall acceptable agreement (Krippendorff's α = .72; Krippendorff, 1980; calculated with the irr package in R; Gamer et al., 2019). After excluding 2,798 records that did not meet the specific selection criteria indicated below, 604 articles remained. During the third step, the first author read the full texts carefully to assess the eligibility of the 604 articles based on the same selection criteria. This led to the exclusion of another 427 articles in the fourth step, leaving 177 articles that meet all selection criteria. Overall, the final data set includes 842 effect sizes (k) from 271 independent samples (m) that were reported in 177 articles published between 1967 and 2019.

To minimize biases induced by selective reporting, we included dissertations, Master theses, and Bachelor theses (10 articles, k = 85). In case a dissertation or thesis was also published in a journal, we included the corresponding journal publication instead. Aside from the search of dissertations and theses, no further steps were taken to identify and collect additional data from unpublished sources. Rather, to identify potential publication bias effects, we relied on methods to assess selective reporting (see the paragraph on selective reporting in the Method and Results sections).

Importantly, we included only articles with accessible full text written in English or German. To be included, the articles had not only to address the relevant topic (i.e., sleep and memory), but also to report at least one memory task that specifically assessed episodic memory. We thus excluded memory tasks such as semantic memory tasks, spatial orientation and navigation tasks, visual discrimination tasks, and prototype learning tasks. Furthermore, we excluded memory tasks focusing on motor components.

Articles were also excluded for three additional reasons. First, to avoid hidden duplicates, we excluded articles that reported relevant results for the current meta-analysis but were taken from previously published articles included in our data set. Second, whenever results from the same study and sample were reported in different articles, we included these articles but coded them as one. For example, one article reported the results of a recognition task for objects (Bennion et al., 2015), whereas recognition results for object backgrounds from the same study were reported in a second article (Bennion et al., 2017) for a subsample of the original sample. In these cases, we coded both articles as one (Bennion et al., 2015) and counted the second article (Bennion et al., 2017) as excluded. Furthermore, several articles had to be excluded because they did not provide sufficient information for effect size computation. In case essential information for the calculation of the effect size was missing in the article, we contacted the authors for this information. If applicable, we also asked the authors for additional missing information (e.g., demographic information) from the same or from other articles. In total, the first author sent 24 emails to request missing information. Five authors (20.83%) provided the requested information. Thus, 20 articles had to be excluded because

the necessary information for effect size computation (e.g., the sample size in the sleep and wake condition) was neither reported in the article nor available upon request. In the next section, we describe the selection criteria in more detail.

Selection Criteria

To be incorporated into the current meta-analysis, articles had to meet the following inclusion criteria: First, the article reports primary empirical results from a sample of healthy drug-free human individuals. Second, the article provides sufficient quantitative data to estimate effect sizes for episodic memory performance in recognition, cued recall, or free recall tasks. Specifically, memory performance is reported as either (a) the percentage of correctly recalled items in a delayed test based on all items learned, (b) the percentage of recalled items in a delayed test based on items correctly recalled in an immediate test right after learning, (c) the corrected hit rate (hit rate - false alarms rate), (d) sensitivity index d' and response bias c based on signal detection theory for recognition tasks, and (e) additional measures of episodic memory such as the reported deviation between learned and retrieved location on a grid for item-context associations. If memory performance was not reported in any of these ways but sufficient information to transform data accordingly was provided, we included the article. For example, we incorporated articles that reported the number of correctly recalled items and the number of tested items in the delayed memory test because the percentage measure required for our analysis can be derived from this information. In contrast, we excluded articles that reported only the aggregated number of correctly identified old and new items in a recognition task (e.g., Maurer et al., 2015), or memory performance in terms of memory gains, memory losses, recovered memories, and memories lost during learning (e.g., Fenn & Hambrick, 2015). Notably, although we included the response bias measure c for exploratory reasons (k = 19), it is excluded from the statistical analyses because it is no memory measure. Third, the article compares memory performance between a sleep and a wake condition. We included any sleep study design as long as sleep versus wakefulness took place within the retention interval. Last but not least, we included an article only if the retention interval

difference between sleep and wake conditions was less than 8 hr. Among those articles that met this criterion, the retention intervals differed between sleep and wake conditions in 14 articles (k = 42) and differences ranged from 36 s to 6.99 hr (M = 2.95 hr, SD = 2.36). Notably, the retention interval was longer for the sleep than for the wake condition in almost all of these studies (k = 38, m = 14). Solely for Prehn-Kristensen et al.'s (2013) adult sample, a longer retention interval (4 hr) was reported for the wake condition (k = 4, m = 1). Because most of the retention intervals are longer in sleep compared to wake conditions, assessing the sleep benefit in episodic memory on the basis of the full data set can be considered as conservative. We therefore performed a sensitivity analysis to check whether the result pattern differs when all effect sizes are excluded for which differences in retention interval lengths exist or no information about the length of retention intervals was reported (see the Sensitivity Analyses section for more details).

Coding of the Moderators

Coding was conducted by the first author. In addition, 45 articles (25%) randomly selected from the 177 articles included in the meta-analysis were also coded by a second coder. The second coder was trained by the first author on an independent training data set of 10 articles for which only the coding of the first author is included in the data set for the present meta-analysis. Discrepancies in the coding were discussed and resolved by the two coders throughout the coding process. Overall, the agreement between the two raters for the moderator variables included in the present meta-analysis was high (.76 \leq Krippendorff's $\alpha \leq 1$). We report Krippendorff's α as an index of the interrater agreement for each moderator variable and provide a descriptive summary of sample, study, and moderator characteristics in Table 1.¹² Next, we describe coding of some of the moderator variables in more detail.

Emotional stimuli differ in valence (e.g., negative, positive) and arousal (e.g., negative stimuli can be more arousing than positive ones). However, because several studies merely reported valence

¹² Note that more variables were coded than included in the current meta-analysis. As these additional variables are not part of the present meta-analysis, we will not provide any further information about them beyond the information in Table 1. For more details on these variables see the codebook in the additional materials provided on the OSF (https://osf.io/dr6hw/?view_only=c55b3737b6c642cfb170276088930a04).

but no arousal data, we based coding of the moderator variable "emotionality" on the valence dimension only. Note, however, that emotionality as coded here is no pure measure of emotional valence because valence and arousal are likely confounded. Specifically, we coded stimuli as negative, positive, or neutral when explicitly stated. If no explicit information about the emotionality of the stimuli was given but we had reasons to assume that the stimuli were neutral (e.g., after checking the provided stimuli), we coded them accordingly. Furthermore, we coded stimuli as emotional when negative and positive stimuli were summarized, or stimuli were specified only as emotional or arousing. If articles reported results, taking negative, positive, and neutral stimuli together, we categorized these effect sizes as mixed.

Because several studies did not report the mean age (k = 359), we defined three discrete age ranges: Children and adolescents under 18 years, adults between 18 and 59 years, and older adults from the age of 60 years. With this definition of the age ranges, we followed Ohayon et al. (2004), who investigated age related changes in sleep parameters meta-analytically.

For stimulus material, we coded any type of verbal study material, including words presented in either language, artificial words, nonsense syllables, definitions, and prose passages as verbal material. In contrast, illustrations of faces, persons, objects, animals, cartoons, and scenes were coded as pictures. Furthermore, we coded the stimulus material as mixed or other, whenever various materials were combined (e.g., combinations of words and pictures), or study materials consisted of other stimuli such as odors, sounds, locations on a grid, or colored picture frames.

To capture differences between sleep study designs, we defined seven categories:

- Natural sleep and wakefulness: Participants in the sleep condition learn the material in the
 evening and are tested after a night of sleep in the morning. Conversely, participants in the
 wake condition learn the material in the morning and are tested after a day of wakefulness in
 the evening.
- Daytime nap: Participants in the sleep condition nap during the day (i.e., before 7 p.m.),
 whereas participants in the wake condition stay awake throughout this time.

- Nighttime nap: Like "daytime nap" with the exception that participants in the sleep condition nap during the night (i.e., after 7 p.m.).
- Total sleep deprivation with recovery night: Participants in the sleep condition sleep normally,
 whereas participants in the wake condition stay awake in the first or second night after
 learning. Memory testing for sleep and wake conditions takes place after one or more recovery
 nights.
- Total sleep deprivation without recovery night: Like "total sleep deprivation with recovery night" with the exception that memory testing for sleep and wake conditions takes place directly after sleep deprivation.
- SWS deprivation: Participants in the sleep condition sleep during the second part of the night to reduce SWS, whereas participants in the wake condition stay awake during this time.
 Alternatively, throughout a full night of sleep, participants in the SWS deprivation condition are awakened every time the first epoch of SWS occurs. This sleep condition is compared with an equally long retention interval filled with daytime wakefulness.
- REM sleep deprivation: Like "SWS deprivation" with the exception that participants in the sleep
 condition sleep during the first part of the night to reduce REM sleep or are awakened every
 time the first epoch of REM sleep occurs throughout a full night of sleep.

We coded accuracy of effect size computation as high whenever we were able to retrieve the information required for effect size computation directly from the text or from tables. In particular, the following information was considered necessary: Means and standard deviations or standard errors or other statistics such as t or F values from which effect sizes can be computed directly. If we could not retrieve this information, we coded accuracy of effect size computation as low. In such cases, we used the open source WebPlotDigitizer (Rohatgi, 2019) to obtain means and standard deviations or standard errors from graphs (k = 232). If this was also not possible (k = 10), we used replacement statistics to estimate the effect size (e.g., frequencies from a two by two frequency table, p value of a t test).

Furthermore, TST, SWS, and REM sleep were captured only for the experimental day or night in which sleep versus wakefulness took place. To illustrate, consider a total sleep deprivation experiment where participants in the sleep condition sleep normally and participants in the wake condition stay awake during the first night after learning. In the second night after learning, participants in both the sleep and wake condition are allowed to sleep before they perform a delayed memory test the next day. Because sleep versus wakefulness differed between the conditions in the first night only, we coded the TST, SWS, and REM sleep for the first night but not for the second night after learning. In addition, whenever TST was not reported, we calculated the difference between the planned start and end time of the sleep interval provided in the experimental protocol and used this difference as TST measure.

Finally, whenever the length of the retention interval was not reported, we used the planned or actual mean start time of the learning and testing session to calculate the retention interval. We decided to use the start time of the learning sessions because it varies less between participants and is reported more often than the end time. Moreover, when the retention interval was reported separately for the sleep and wake condition, we computed the mean interval length (unless the difference between conditions was 8 hr or more, see the Selection Criteria section). In case the retention interval was reported only for one of the conditions, we assumed that this interval represents both conditions. Table S4 in the supplemental materials provides correlations among moderator variables included in the present meta-analysis.

Effect Size Computation

To obtain unbiased effect size estimates also for small sample sizes, we used Hedges' g as the effect size measure. As such, Hedges' g measures the standardized difference between the group means and is asymptotically equivalent to Cohen's d (Cohen, 1988). However, Hedges' g is more conservative than Cohen's d because it corrects for the positive bias of Cohen's d in small samples (Borenstein et al., 2009; Ellis, 2010). This notwithstanding, Hedges' g and Cohen's d can be interpreted in the same way: Effect sizes around 0.20, 0.50, and 0.80 are usually considered as small,

medium, and large effects, respectively. However, this classification should only be seen as a rough guideline that may require adaption to the relevant field of research when evaluating the size of an effect (see Cohen, 1988). We estimated Hedges' g by first determining Cohen's d (Step 1) and then deriving Hedges' g from Cohen's d (Step 2). These two steps are described in more detail below.

In Step 1, we calculated Cohen's d on the basis of the reported means and standard deviations or standard errors. If this information was not available, we used the value of a t test, F test, or related statistics (e.g., the p value of a t test) to derive Cohen's d. The effect size estimates and their respective sampling variance were calculated with the web based effect size calculator provided by David B. Wilson¹³ (Lipsey & Wilson, 2001).

To compute Cohen's *d* for within-subject designs, we used the standard deviation of the wake condition instead of the pooled standard deviation of the sleep and wake condition as denominator (Becker, 1988; Morris & DeShon, 2002). This approach has two advantages: First, Cohen's *d* can be calculated without knowing the correlation between the sleep and wake condition. Second, Cohen's *d* can be interpreted in the same way for within and between designs. That is, both measures inform about the mean difference between sleep and wake conditions when the within-group standard deviation is used as the measurement unit. Assuming homogeneity of variances, estimating this standard deviation from a single group (within-subject design) or from both groups using a pooled estimate (between-subjects design) should lead to very similar results because these estimates will differ as a consequence of sampling error only. However, because sleep may affect the standard deviation in sleep conditions in unknown ways, we used the standard deviation of the wake condition to standardize the mean differences in within-subject designs. This results in effect size estimates that are comparable between (a) within-subject studies with unknown and perhaps varying correlations between sleep and wake conditions, and (b) between-subjects studies.

In the current meta-analysis, we used inverse-variance weighting which is based on the sampling variances of the effect sizes. To compute the sampling variance of an effect size in within-subject

¹³ https://campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php

designs, the correlation between the sleep and wake condition is necessary. As none of the included articles reported correlations between sleep and wake conditions, we estimated the correlation with the help of the t value whenever possible (k = 64). In case we could not retrieve the t value, we used the corresponding p value to get an estimate of the t value. However, if we could not obtain the correlation as just described, we used the mean correlation of the existing data to estimate the sampling variance of an effect size for within-subject designs (k = 91).

In Step 2, we converted Cohen's d into Hedges' g in R (Version 4.0.0) by using the correction factor J (see formula 4.22 in Borenstein et al., 2009; Hedges, 1981). Importantly, we defined all effect sizes such that positive values always indicate a sleep benefit in episodic memory and negative values a sleep disadvantage compared to wakefulness.

Statistical Analysis

Articles often report several effect sizes of interest on the basis of the same—or partly the same—participant sample. For example, an article might report both cued recall and recognition test results for the same sample of participants, with the recognition task conducted after the cued recall task. Both effect sizes are of interest but originate from the same sample. In another article, a daytime wake condition and a total sleep deprivation condition are compared to the same sleep condition. Again, both effect sizes are relevant, but they are based on overlapping samples. In either case, the independence assumptions of standard meta-analyses are violated.

To account for dependencies among effect sizes, we applied multilevel meta-analysis with robust variance estimation (RVE) and clustered at the level of samples. Specifically, like in a classical random-effects model, we allowed effect sizes to vary systematically beyond sampling error. That is, we took the sampling variances of effect sizes at the sample level into account (Level 1) and allowed different effect sizes to vary within the same sample (e.g., variations due to a within-subject manipulation of the retrieval procedure or emotionality of study materials; Level 2), and additionally

¹⁴ We estimated the correlation by $r = (SD_{\text{wake}}^2 + SD_{\text{sleep}}^2 - SD_{\text{diff}}^2)/(2 * SD_{\text{wake}} * SD_{\text{sleep}})$, with $SD_{\text{diff}}^2 = [n * (M_{\text{sleep}} - M_{\text{wake}})^2]/t^2$ (Morris & DeShon, 2002). Here, t denotes the t-test statistic from a repeated-measures t test and n indicates the number of participants who completed both the sleep and wake condition.

between different samples (e.g., variations based on sleep study designs or age groups; Level 3). The three-level model used here extends the classic random-effects model to account for the hierarchical structure of the data. However, to take this data structure adequately into account, knowledge of the covariances between effect sizes is also required. Because this information is commonly unknown, we used RVE to correct for dependencies among effect sizes (Hedges et al., 2010).

All analyses were conducted using the R packages metafor (Viechtbauer, 2010) and clubSandwich (Pustejovsky, 2020). In a first step the three-level model was fitted with metafor by using the restricted maximum likelihood estimation method (REML) and inverse-variance weighting. This typical weighting scheme assigns more weight to more precise effect sizes (i.e., effect sizes with lower sampling variance). In a second step we used the RVE method implemented in clubSandwich to correct for dependencies among effect sizes. Importantly, the Type I error rates for RVE-based test statistics and confidence intervals can be inflated for covariates that are unbalanced or have high leverage values. This holds specifically if the number of independent samples is small or moderate (Tipton, 2015; Tipton & Pustejovsky, 2015). An example for an unbalanced covariate would be the potential moderator "definition of dependent variable" with 83.96% of the effect sizes falling in one category and with only 16.04% in the other. Furthermore, consider the variable "retention interval" as an example for a high leverage covariate: The majority of retention intervals are between 45 min and 2 days long, but there are also four cases with a duration of 30 days.

To provide for Type I error rates close to the nominal level of α = .05, we applied the generally recommended small sample corrections that includes a Satterthwaite approximation for the degrees of freedom (Tipton, 2015; Tipton & Pustejovsky, 2015). Note that, as a consequence of the small sample corrections, the degrees of freedom depend not only on the number of independent samples, but also on characteristics of the covariates (e.g., the presence of high leverage values). Therefore, the degrees of freedom and thus the power of the statistical tests (i.e., small-sample corrected *F* and *t* tests) differ between moderator variables and also between their respective levels.

Simulation studies have shown that small sample corrections for t tests work best if there are more than four degrees of freedom (Tipton, 2015). Otherwise, p values tend to behave anticonservatively and thus cannot be trusted. Consequently, a more rigorous significance level than α = .05 should be adopted in those cases (Tanner-Smith et al., 2016). We therefore used a three-level meta-regression model with small-sample corrected RVE and a significance level of α = .05 in general, but switched to a significance level of α = .01 whenever $df \le 4$.

We provide heterogeneity statistics in the notes of Table 2. Note that the reported statistics and tests designed to quantify effect size heterogeneity might be biased and should thus be interpreted with caution. Because the covariance between the effect sizes is unknown and RVE is a problematic method to quantify the heterogeneity within and between the samples, it is hardly possible to obtain reliable heterogeneity statistics and tests. However, this is no major problem for the current meta-analysis as we focus on the average effect of sleep on episodic memory and, in addition, the impact of potential moderator variables. For these goals, the RVE method is known to be particularly useful (Tanner-Smith et al., 2016; Tipton, 2015) and thus the method of choice in our case.

Note also that we used studentized residuals and Cook's distance to detect outlier and influential cases in the three-level model without RVE (Viechtbauer & Cheung, 2010). To our knowledge, outlier and influence diagnostic methods are not yet available for meta-analysis models with RVE. However, we trust that outlier diagnostics are not severely influenced by using or not using RVE. In the following section, we describe our analytic strategy, the sensitivity analyses performed, and the checks for selective reporting.

Analytic Strategy

To estimate the overall effect of sleep on episodic memory, we fitted an intercept-only three-level model with RVE. We then carried out the following five steps to investigate potential moderators: First, we omitted all missing values on the moderator variables using list-wise deletion. Second, we checked the number of effect sizes per moderator level for each categorical moderator. Because four or fewer effect sizes cause problems for small sample corrected *t* tests in terms of

power (Tanner-Smith et al., 2016; Tipton, 2015), we excluded all moderator levels with four effect sizes or less. Third, we analyzed the impact of each moderator in the context of all other moderators using a multiple-moderator model with RVE. Specifically, we entered various dummy coded categorical variables and untransformed continuous variables into the same meta-regression. This approach allowed us to investigate the impact of each moderator while controlling for all other moderators statistically. To test whether the sleep benefit in episodic memory varies across the levels of a specific moderator, we used small-sample corrected F tests (i.e., approximate Hotelling's T^2 [AHT] F tests; Tipton & Pustejovsky, 2015). Moreover, we used small-sample corrected T tests to examine whether the effect of sleep on episodic memory differs significantly from zero for a specific moderator level (Tipton, 2015). Fourth, to report and visualize the results of the moderator analyses, we applied the moving constant technique (Johnson & Huedo-Medina, 2011). Note that effect sizes reported for moderator levels are marginal average effects after adjusting for variation in the full set of potential moderators included in the respective multiple-moderator model. Fifth, to more directly test the hypotheses derived from consolidation or interference theories for potential moderators, we used Bayesian evaluation of informative hypotheses.

As such, informative hypotheses are statements derived from theory about directional differences and equalities between parameters of interest (Hoijtink et al., 2019; van Lissa et al., 2021). In the following, we will denote informative hypotheses derived from consolidation theories by H_c and informative hypotheses derived from interference theories by H_i . For focal and parity check variables, these informative hypotheses are compared to a null hypothesis H_0 (i.e., the moderator variable has no effect), and an unconstrained hypothesis H_u (i.e., the moderator variable can have an effect in any direction or no effect at all). Because H_u places no constraints on the parameters of interest, it includes all possible orderings of the parameters and is used as a fail-safe hypothesis in case none of the specified informative hypotheses are supported by the data (Hoijtink et al., 2019; van Lissa et al., 2021). For exploratory and control variables, we simply compared H_0 and H_u as we did not have any clear-cut predictions. Bayesian informative hypothesis evaluation was

performed with the R package bain (Gu et al., 2020). For categorical moderator variables, the hypotheses referred to Hedges' g across the levels of a moderator (i.e., multiple group analysis). For continuous moderator variables, the hypotheses referred to the regression coefficient β (i.e., single group analysis). All necessary information to test informative hypotheses are provided in Table 2 and the supplemental materials (i.e., the number of effect sizes, k; Hedges' g, β , and respective SEs).

The results of Bayesian hypothesis tests—Bayes factors (BF) and posterior model probabilities (PMP)¹⁵—are sensitive to the fraction of information in the data that is used to specify the variance of the prior. By default, the bain package uses a minimal fraction of the data to compute the variance of the prior distribution (Gu et al., 2020; Hoijtink et al., 2019). This results in the largest prior variance and thus gives H_0 the largest support before observing the data (Hoijtink et al., 2019). Considering possible selective reporting in the literature, it seems reasonable to follow such a conservative test strategy. Nevertheless, using only the minimal fraction of information might be too conservative. Therefore, we additionally conducted sensitivity analyses in which we used twice and three times the minimal fraction to specify the prior variance. Notably, although BFs and PMPs are sensitive to the fraction of information used, the hypothesis that received most support from the data did not change for all moderator variables except for emotionality and feedback. In the Results section, we therefore report the BF comparing an informative hypothesis with its complement¹⁶ (BF.a) and PMPs for the focal and parity check variables using the minimal fraction. For the moderator variables "emotionality" and "feedback" we additionally report BF_{-a} and PMP of the sensitivity analysis using three times the minimal fraction, for which the data favored another hypothesis. Table S62 in the supplemental materials provides the results for all corresponding sensitivity analyses and for each variable included in the main moderator analysis.

¹⁵ The posterior model probabilities measure the support of a hypothesis on a scale from 0 to 1. They can be used to select from the set of considered hypotheses the hypothesis that receives the most support from the observed data (Hoijtink et al., 2019; van Lissa et al., 2021).

 $^{^{16}}$ The complement is an alternative hypothesis (H_a) that includes all possible orderings of the parameters of interest except the ordering specified by the informative hypothesis under consideration. Comparing an informative hypothesis with its complement provides information about how much the observed data supports the informative hypothesis (Hoijtink et al., 2019; van Lissa et al., 2021).

Because (a) sleep parameters were not always assessed or reported and (b) TST, SWS, and REM sleep are strongly correlated by default ($r_{TST,SWS} = .85$, $r_{TST,REM} = .96$, $r_{SWS,REM} = .81$, all ps < .001), we performed three subgroup moderator analyses for TST, SWS, and REM sleep in addition to the main moderator analysis for the total data set. The following 15 moderator variables were included in the main moderator analysis: emotionality (5 levels), restudy (2 levels), testing (2 levels), feedback (2 levels), encoding instruction (2 levels), directed forgetting instruction (3 levels), retrieval procedure (3 levels), age (3 levels), definition of dependent variable (2 levels), stimulus material (3 levels), sleep study design (7 levels), experimental design (2 levels), randomization (2 levels), accuracy of effect size computation (2 levels), and retention interval. In total, this analysis was based on 777 effect sizes (46 effect sizes were excluded due to missing values on the moderator variables). Because all moderator levels contained more than four effect sizes, no moderator or moderator level had to be excluded. In the TST, SWS, and REM sleep subgroup analyses, we entered the same variables as in the main moderator analysis but included TST, SWS, or REM sleep as an additional continuous moderator, respectively. Overall, 641 effect sizes were included in the TST subgroup analysis, and 427 effect sizes each in the subgroup analyses for SWS and REM sleep (effect sizes excluded due to missing values: $k_{TST} = 182$, $k_{SWS} = k_{REM} = 396$). Note that for the TST, SWS, and REM sleep subgroup analyses, one category of the moderator variable "emotionality" (emotional) had to be excluded because it contained only three effect sizes or no effect sizes at all.

Furthermore, we carried out an additional moderator analysis on a subset of the data containing only single words and word pairs presented in any language, including artificial words and nonsense syllables (k = 335). The purpose of this analysis is to provide more specific and detailed information on often used verbal learning material. Single words and word pairs have been targets of experimental research on episodic memory for more than a century (e.g., Ebbinghaus, 1885; Müller & Pilzecker, 1900) and continue to be used frequently in research on sleep and memory until today (e.g., Diekelmann et al., 2009; Klinzing et al., 2019). Note that we excluded the moderator variable "emotionality" in this subgroup analysis because only one category (neutral) contained more than

one effect size. In addition, the moderator "stimulus material" had to be removed because single words and word pairs were considered only. Apart from these two exceptions, all moderators and moderator levels of the main analysis were included (k = 319).

Sensitivity Analyses

To check the robustness of our results we performed several additional sensitivity analyses. Specifically, we conducted seven sensitivity analyses on the main data set and the four subgroup data sets for TST, SWS, REM sleep, and single words and word pairs. First, we examined whether the result pattern differs when all effect sizes are excluded for which the correlation between sleep and wake conditions in within-subject designs was unknown (k = 91). As noted before, whenever we could not obtain the correlation, we used the mean correlation of the existing data to estimate the sampling variance. Because this may distort the results, it is important to check whether exclusion of these effect sizes affects our findings. Second, we checked whether the result pattern differs when all effect sizes are excluded for which memory performance was assessed with other measures than the percentage of correctly recalled items, the corrected hit rate, or d' (k = 46). In a third sensitivity analysis, we tested the robustness of the result pattern regarding experimentally introduced interference. In this analysis, we excluded all effect sizes for which proactive interference was introduced at encoding (k = 6) or retroactive interference was introduced at encoding (k = 7) or retrieval (k = 53). Fourth, we examined whether the result pattern differs when all effect sizes are excluded for which TMR was applied during sleep (k = 7) or wakefulness (k = 2) or during sleep and wakefulness (k = 19). Fifth, we tested whether excluding all effect sizes of sleep study designs that involve some form of sleep deprivation (k = 235) affects the results. For this analysis, we excluded all sleep study designs except (a) natural sleep and wakefulness, and (b) daytime naps. Sixth, to check the robustness of our results with regard to differences in retention intervals between sleep and wake conditions, we excluded all effect sizes for which the length of the retention interval differed between conditions (k = 42) or was not reported (k = 10). Finally, we examined the robustness of the result pattern regarding influential outliers (k = 3). We observed five effect sizes with studentized

residuals larger than 1.96. To determine whether these outliers are also influential, we calculated Cook's distance. This resulted in the identification and removal of three influential outliers.¹⁷

In addition, we conducted three sensitivity analyses on the main data set only, one sensitivity analysis on the subgroup data sets for TST, SWS, and REM sleep only, and another sensitivity analysis on the subgroup data set for TST only. The first sensitivity analysis on the main data set checked whether the result pattern differs when assuming intentional learning for effect sizes with missing values on the focal variable "encoding instruction" (k = 29). In typical experiments, intentional learning appears more likely than incidental learning, as participants might expect that their memory for studied items will be tested. Therefore, we coded missing values in encoding instruction as intentional learning in this sensitivity analysis. In a second sensitivity analysis on the main data set, we included the publication year as an additional control variable to assess possible differences in result patterns between past and recent research.¹⁸ The third sensitivity analysis on the main data set tested the robustness of our results with respect to the assumed value p for sampling correlation. As noted before, we applied RVE to correct for dependencies among effect sizes because this method does not require knowledge of the exact correlations between effect size estimates (Pustejovsky & Tipton, 2021; Rodgers & Pustejovsky, 2020). For the RVE, we used a threelevel model as working model that assumes no correlation in sampling errors for effect sizes based on the same—or partly the same—participant sample ($\rho = .00$). Although the exact value of the assumed sampling correlation is typically not crucial for the results obtained (Rodgers & Pustejovsky, 2020), we varied p from .10 to .90 in steps of .10. Thus, we conducted nine additional sensitivity analyses for the overall effect and the main multiple-moderator analysis.

In the sensitivity analysis on the subgroup data sets for TST, SWS, and REM sleep only, we included sleep parameter assessment as additional exploratory variable to check whether our results depend on the assessment method (e.g., self-report, PSG). Specifically, we captured sleep parameter

¹⁷ Two influential outliers were identified in Mawdsley et al. (2014) and a third in Sheth et al. (2012).

¹⁸ A detailed analysis of the relationship between publication year and (a) effect sizes, (b) sample sizes, and (c) number of missing values in the main data set is provided in the supplemental materials.

assessment with four categories: (a) self-report, (b) sleep tracker (e.g., actigraphy, pulse monitoring), (c) PSG at home, and (d) PSG in laboratory. Finally, we conducted an additional sensitivity analysis on the TST subgroup data to explore whether there is a quadratic instead of a linear relation between TST and the sleep benefit in episodic memory. In total, we addressed 12 different methodological issues by performing sensitivity analyses on the main and subgroup data sets. For each of the sensitivity analyses, we removed moderator levels with four or less effect sizes, which in turn could lead to the exclusion of moderators (for the included moderators and moderator levels see Tables S12 to S61 in the supplemental materials).

Selective Reporting

The validity of meta-analyses can be jeopardized by selective reporting of statistically significant findings. Selective reporting has various causes, including publication bias and questionable research practices such as *p*-hacking (e.g., Coburn & Vevea, 2015; Erdfelder & Heck, 2019; John et al., 2012; Simmons et al., 2011; Ulrich et al., 2018; van der Steen et al., 2018). In meta-analyses, selective reporting of significant results in the hypothesized direction can lead to an overestimation of the effect under investigation. Therefore, it is important to examine the presence of selective reporting and to assess its impact on the overall effect size estimate. A traditional method to check for selective reporting is the visual inspection of a funnel plot. In essence, the funnel plot is a scatter plot of the effect size estimates against a measure of precision (Sterne & Egger, 2001). Here, we plotted the effect size estimates against a modified measure of effect size precision¹⁹ to check for asymmetry in the funnel plot.

In addition, we tested and corrected for funnel plot asymmetry by using a modified variant of Egger's linear regression (Egger et al., 1997). This modified test avoids inflated Type I error rates due to correlated effect sizes and due to correlations between effect sizes and their respective standard errors (Pustejovsky & Rodgers, 2019; Rodgers & Pustejovsky, 2020). Thus, we fitted a three-level model with small-sample corrected RVE and included the modified measure of effect size precision

¹⁹ We used a function of the sample size as effect size precision measure, calculated as $[(n_{\text{sleep}} + n_{\text{wake}})/(n_{\text{sleep}} * n_{\text{wake}})]^{1/2}$ for between-subjects designs and $1/N^{1/2}$ for within-subject designs (Pustejovsky & Rodgers, 2019).

as a covariate in the meta-regression. Note that the intercept of this meta-regression can be interpreted as an asymmetry-corrected overall effect size estimate and the slope as a measure of asymmetry. By implication, a statistically significant slope signals substantial asymmetry in the funnel plot. Notably, selective reporting is only one explanation for funnel plot asymmetry. There are also other possible reasons such as extreme heterogeneity between effect sizes due to different study designs. Hence, the reported checks for funnel plot asymmetry do not conclusively test for selective reporting (Pustejovsky & Rodgers, 2019; Rodgers & Pustejovsky, 2020; Sterne & Egger, 2001).

To address this issue, we used the three-parameter selection model (3PSM; Vevea & Hedges, 1995) to check for selective reporting more directly. The 3PSM comprises a model of the data generation process and a model of the selection process. Using maximum likelihood methods, three parameters are estimated: (a) heterogeneity between effect sizes, τ^2 ; (b) corrected overall effect size estimate, β ; and (c) probability ratio of observing a published nonsignificant result relative to a significant one, ψ . In addition, a likelihood ratio test can be used to test for selective reporting. A significant likelihood ratio test statistic indicates that selective reporting is present in the data (Vevea & Hedges, 1995). We chose the 3PSM to test for selective reporting, because—compared to other methods for detecting selective reporting (e.g., p-curve approach, trim-and-fill method, rank correlation test)—this model performed best in recent simulation studies (Carter et al., 2019; McShane et al., 2016; Pustejovsky & Rodgers, 2019; Rodgers & Pustejovsky, 2020). However, the 3PSM can only be applied to data sets with independent effect sizes. We therefore created 20 subsets based on independent effect sizes only by randomly selecting a single effect size per sample. We used the R package weightr (Coburn & Vevea, 2019) to estimate the three parameters for each of the 20 subsets.

Results

Overall, the analyzed data contained 823 effect sizes from 271 independent samples that were reported in 177 articles published between 1967 and 2019 (*Mdn* = 2013). Of the 823 effect sizes,

10.20% came from dissertations and other theses (10 articles, k = 84). Most included articles refer to countries with a western or western-influenced culture: The US takes the lead with 58 articles, closely followed by Germany with 55 articles. By contrast, only a few articles originated from eastern countries (e.g., one article from China and two articles from Japan). Overall, the median sample size was 30 participants, mostly young adults of about 22 years of age with a qualification for college/university entrance or at least 12 years of education. Across all samples, an approximately equal proportion of females and males were studied ($Mdn_{\text{female ratio}}$ = 0.56). For more detailed information on sample, study, and moderator characteristics, see Table 1 and Tables S5 to S7 in the supplemental materials.

In the following section, we first provide overall effect size estimates for the sleep benefit in episodic memory. Next, we report the results from multiple-moderator models to systematically analyze the impact of each moderator while controlling for effects of all other moderators. Finally, we provide checks of and corrections for selective reporting. For all analyses, we report the number of effect sizes (*k*) and the number of independent samples (*m*) that entered into the analysis.

Overall Effect of Sleep on Episodic Memory

One aim of the current meta-analysis is to quantify the overall sleep benefit in episodic memory. For this purpose, we fitted an intercept-only three-level model with RVE to estimate the overall effect of sleep on episodic memory. We observed a moderate sleep benefit in episodic memory for the complete data set, g = 0.44, SE = 0.03, 95% CI [0.39, 0.50], t(252.26) = 16.94, p < .001, k = 823, m = 271. The overall effect and all effect size estimates included in this meta-analysis are depicted in Figure 3 (for more details, see Figure S1 in the supplemental materials). Varying the assumed sample correlation ρ from .10 to .90 when calculating RVEs results in overall effects ranging from g = 0.44 for $\rho = .10$ to g = 0.41 for $\rho = .90$ (see Tables S49 to S57 in the supplemental materials). Hence, results are robust against different assumptions concerning ρ . Moreover, the overall effect also proved to be stable in the other sensitivity analyses. It ranged from g = 0.43 in the sensitivity analysis on

interference introduction to g = 0.48 in the sensitivity analysis on sleep deprivation (see Tables S12, S17, S22, S27, S32, S37, and S42).

Moderator Analyses

To conduct the moderator analyses described above, we fitted several multiple-moderator models with RVE (for details see the Analytic Strategy section). The results of the main moderator analysis are summarized in Table 2. Regarding the TST, SWS, and REM sleep subgroup analyses, we report only the results of the respective sleep parameter. The results for all moderator variables included in the TST, SWS, and REM sleep subgroup analyses are provided in the supplemental materials (Tables S9 to S11). In addition, Table S8 in the supplemental materials provides the results of the subgroup moderator analyses for single words and word pairs. Note that in all moderator analysis tables, the results for categorical moderators are presented in the table body and the results for continuous moderators are reported in the table notes. Note also that the analyzed data subsets differ in the number of effect sizes and thus in the number of moderator variables and moderator levels included. Consequently, the results of the subgroup analyses can differ from the results of the main moderator analyses reported here. Descriptively, however, the observed result patterns remained largely the same across all moderator analyses conducted.

In the following paragraphs, we will report results for different moderator variables and their match with predictions from consolidation and interference theories in a step-by-step fashion. A comprehensive and integrative discussion of the entire pattern of results across moderators will follow in the Discussion section.

Focal Variables

Emotionality. A potential moderator of the sleep benefit in episodic memory is emotionality. According to consolidation theories, the sleep benefit should be largest for negative, followed by positive, and lastly neutral memories. In the current meta-analysis, we added two more categories for exploratory reasons (i.e., the categories "emotional" and "mixed" as described in the Method section). For the main moderator analysis, the AHT *F* test indicated no statistically significant

differences between the five moderator levels, F(5, 10.71) = 1.66, p = .231, k = 777, m = 254. However, there was a statistically significant sleep benefit for neutral, q = 0.40, t(145.57) = 14.64, p <.001, negative, g = 0.58, t(36.19) = 8.33, p < .001, and positive memories, g = 0.45, t(8.98) = 6.95, p = 0.95.001. A descriptive pattern in the direction predicted by consolidation theories is discernible, with the largest sleep benefit for negative, followed by positive, and lastly neutral memories (see Figure 4A). Note that the number of effect sizes and the associated number of degrees of freedom for emotional memories is quite low (k = 9, m = 6, df = 4.76) and for mixed memories too small for a meaningful interpretation (k = 11, m = 6, df = 3.91). Informative hypothesis evaluation showed that both H_0 (BF_{0a} = 7.22) and H_c (BF_{ca} = 7.80) received support from the data with H_0 being the hypothesis favored most (PMP = .64). However, when using three times the minimal fraction, H_c received most support from the data (PMP = .48; BF_{ca} = 7.69, BF_{0a} = 2.41), indicating that the results are sensitive to the fraction of information used to specify the prior variance. These results show that H_c is a plausible hypothesis that should not be ignored. Moreover, except for the SWS and REM sleep subgroup and respective sensitivity analyses that showed a reverse order for positive and neutral memories, all analyses performed resulted in the same descriptive pattern as the main moderator analysis. To conclude, in line with consolidation theories we found a descriptively somewhat larger sleep benefit for negative information compared to positive and neutral information, a difference that was, however, not statistically significant.

Restudy. According to interference theories, the sleep benefit should be larger for restudy than for no-restudy conditions. In fact, we observed statistically significant differences between the two levels of the moderator "restudy", F(2, 88.37) = 9.71, p = .002, k = 777, m = 254. As predicted, the sleep benefit was larger for restudy, g = 0.55, t(81.00) = 12.07, p < .001, than for no-restudy conditions, g = 0.38, t(125.59) = 12.72, p < .001 (see Figure 4B). Correspondingly, informative hypothesis evaluation showed that H_i received most support from the data (PMP = .62; BF_{ia} = 899.89, BF_{0a} = 0.23). Furthermore, the result pattern of the main moderator analysis was evident in all performed sensitivity analyses, demonstrating its robustness. In a nutshell, restudy compared to no

restudy appears to boost the sleep benefit in episodic memory, as predicted by interference theories.

Testing. Interference theories predict no substantial difference in the sleep benefit between encoding conditions with and without testing. Supporting this prediction, the results showed no statistically significant differences in the sleep benefit between conditions, F(2, 58.70) = 1.32, p = .256, k = 777, m = 254; without testing: g = 0.46, t(90.40) = 10.61, p < .001, k = 303, m = 98; with testing: g = 0.40, t(94.03) = 12.03, p < .001, k = 474, m = 178 (see Figure 4C). This finding was confirmed by informative hypothesis evaluation showing that H_i (which is equivalent to H_0 in this case) received most support from the data (PMP = .93; BF_{ia} = 13.48). Moreover, across all analyses performed, the same result pattern as in the main moderator analysis emerged. Hence, as expected in the framework of interference theories, there is no evidence that testing moderates the sleep benefit in episodic memory.

Feedback. According to interference theories, the sleep benefit should be larger in feedback than in no-feedback conditions. We found no statistically significant differences between the two levels of the moderator "feedback", F(2, 91.55) = 3.49, p = .065, k = 777, m = 254. Descriptively, however, the sleep benefit was larger for feedback, g = 0.52, t(99.33) = 8.79, p < .001, than for no-feedback conditions, g = 0.39, t(119.45) = 12.57, p < .001 (see Figure 4D). In line with this, informative hypothesis evaluation suggests that H_i is a plausible hypothesis that should not be ignored. Specifically, using the minimal fraction, the results showed that both H_0 (BF_{0a} = 3.21) and H_i (BF_{ia} = 42.49) received support from the data with H_0 being most supported (PMP = .52). In the sensitivity analysis using three times the minimal fraction, however, H_i received most support from the data (PMP = .41; BF_{ia} = 42.49, BF_{0a} = 1.85). Moreover, as in the previous analyses, the descriptive pattern of the main moderator analysis was mirrored by all other analyses performed for the "feedback" moderator. Overall, as predicted by interference theories, feedback at encoding resulted in a larger sleep benefit than no feedback; this effect, however, was not statistically significant.

Encoding Instruction. Learning can be intentional or incidental. As outlined before, interference theories predict a larger sleep benefit for intentional than for incidental learning. As indicated by the AHT F test, the two levels of the moderator "encoding instruction" did not differ significantly, F(2, 49.52) = 1.02, p = .318, k = 777, m = 254. In contrast to the prediction, the sleep benefit appeared to be descriptively larger for incidental learning, g = 0.49, t(50.89) = 6.35, p < .001, than for intentional learning, g = 0.40, t(110.38) = 12.42, p < .001 (see Figure 4E). Informative hypothesis evaluation showed that H_0 received most support from the data (PMP = .91; BF_{ia} = 0.15, BF_{0a} = 12.05). Except for the TST subgroup sensitivity analysis that included sleep parameter assessment as exploratory variable, all performed analyses showed the same descriptive outcome as the main moderator analysis. In sum, results for intentional versus incidental learning are not easily reconciled with interference theories.

Directed Forgetting Instruction. Consolidation theories predict a larger sleep benefit for information instructed to be remembered than for information instructed to be forgotten. For exploratory reasons, we added a third moderator level to capture the situation where no instructions were given. We found no statistically significant differences between the three moderator levels, F(3, 10.59) = 0.82, p = .467, k = 777, m = 254. However, the sleep benefit was descriptively larger when no instructions or remember instructions were given; no instruction: g = 0.43, t(147.43) = 15.50, p < .001; remember instruction: g = 0.50, t(11.83) = 4.54, p = .001; forget instruction: g = 0.30, t(13.16) = 2.31, p = .038 (see Figure 4F). Informative hypothesis evaluation showed that H_0 received most support from the data (PMP = .58; BF_{0a} = 3.87). However, there was also some support for H_c (BF_{ca} = 7.33). Notably, all performed analyses showed a descriptively larger sleep benefit following remember than forget instructions. In summary, observed effects were in line with consolidation theories but not statistically significant.

Retrieval Procedure. Whereas consolidation theories predict no substantial differences in the sleep benefit between retrieval procedures, interference theories predict the largest sleep benefit for free recall, followed by cued recall, and lastly recognition tasks. In the main moderator analysis,

the AHT F test of differences between the three retrieval procedures was not statistically significant, F(3, 38.35) = 1.39, p = .261, k = 777, m = 254. We found statistically significant sleep benefits on each of the three moderator levels. In addition, the results showed a descriptive pattern in the direction predicted by interference theories, with the largest sleep benefit for free recall, q = 0.49, t(41.23) =7.62, p < .001, followed by cued recall, g = 0.45, t(63.60) = 10.10, p < .001, and recognition, g = 0.38, t(62.78) = 9.30, p < .001 (see Figure 4G). Note that the moderator levels "cued recall" (k = 336) and "recognition" (k = 347) contain more than 3 times as many effect sizes as moderator level "free recall" (k = 94). Because free recall tasks are more commonly used for verbal than for other study materials (e.g., pictures, odors), the effect sizes are more evenly distributed across the moderator levels in the subset restricted to verbal learning materials (i.e., single words and word pairs). In fact, the result pattern of the main moderator analysis was even more pronounced in the subgroup analysis for single words and word pairs, with statistically significant differences between the three moderator levels, F(3, 28.05) = 5.06, p = .013, k = 319, m = 135; free recall: g = 0.63, t(38.66) = 8.45, p = 0.63< .001; cued recall: g = 0.47, t(38.13) = 8.49, p < .001; recognition: g = 0.33, t(19.27) = 3.83, p = .001(see Table S8). In the main data set, informative hypothesis evaluation showed that H_c (equivalent to H_0) received most support from the data (PMP = .95; BF_{ca} = 89.92, BF_{ia} = 7.62). In the subset of single words and word pairs, however, H_i received the most support (PMP = .50; BF_{ca} = 4.62, BF_{ia} = 35.01). We found the descriptive result pattern of the main moderator analysis in all performed analyses, except for the sensitivity analysis on memory performance measures where the sleep benefit was largest for cued recall (see Table S17). Thus, retrieval procedure appears to moderate the sleep benefit in episodic memory, at least when single words and word pairs are used as study materials. In sum, the results provide initial support for the prediction of interference theories that the sleep benefit is largest in free recall, followed by cued recall, and lastly recognition tasks.

Parity Check Variables

Age. Both consolidation and interference theories predict that the sleep benefit in episodic memory decreases with advancing age. Thus, the sleep benefit should be largest for children and

adolescents, followed by adults, and older adults. The main moderator analysis, however, showed no statistically significant difference in the sleep benefit between children and adolescents, adults, and older adults, F(3, 20.24) = 0.53, p = .600, k = 777, m = 254. Although no statistically significant differences between age groups emerged, children, adolescents and adults benefited more from sleep than older adults at the descriptive level; children and adolescents: g = 0.41, t(24.96) = 3.69, p = .001; adults: g = 0.43, t(145.96) = 16.16, p < .001; older adults: g = 0.26, t(8.98) = 1.70, p = .124 (see Figure 4H). Informative hypothesis evaluation showed that H_0 received most support from the data (PMP = .94; BF_{ca} = BF_{ia} = 2.93, BF_{0a} = 47.30). In contrast to the main moderator analysis, some sensitivity and subgroup analyses showed the largest descriptive sleep benefits for children and adolescents or indicated similar descriptive sleep benefits for the moderator levels "children and adolescents" and "adults" (see, for example, Tables S9, S12, S22, S42, and S53 to S57). Yet, all performed analyses showed the lowest descriptive sleep benefit for older adults. Thus, whereas our results did not confirm a decreasing sleep benefit across the initial and middle parts of the life span, we observed descriptively smaller sleep benefits in episodic memory for older adults. This is roughly in line with the prediction of consolidation and interference theories.

Definition of Dependent Variable. Next, we investigated whether the sleep benefit is larger for direct measures of retention (i.e., the performance difference immediate test - delayed test) than for memory measures using delayed tests only. As hypothesized, the difference between the two definitions for the dependent variable was statistically significant, F(2, 41.18) = 24.87, p < .001, k = 777, m = 254. Specifically, there was a statistically significant sleep benefit for both moderator levels, but the sleep benefit was larger for direct measures of retention, g = 0.58, t(58.47) = 14.06, p < .001, than for delayed test memory performance, g = 0.39, t(148.39) = 14.88, p < .001 (see Figure 4I). This hypothesis also received the most support in Bayesian informative hypothesis evaluation (PMP = .66; BF_{ca} = BF_{ia} =23955.90, BF_{0a} = 0.01). Across all analyses performed, the same descriptive result pattern as in the main moderator analysis emerged for this moderator variable. In conclusion, the size of the sleep benefit in episodic memory depends on the definition of the dependent variable; it is more

pronounced for direct measures of retention than for measures of delayed test memory performance.

TST. Both consolidation and interference theories predict an increase of the sleep benefit with longer TST. However, the results from the TST subgroup analysis showed no statistically significant effect of TST on the sleep benefit, β = -0.003, SE = 0.03, 95% CI [-0.07, 0.07], t(14.68) = 0.09, p = .931, k = 641, m = 207 (see Table S9). This finding was mirrored by informative hypothesis evaluation: H_0 received most support from the data (PMP = .93; BF_{ca} = BF_{ia} = 0.86, BF_{0a} = 25.21). Across all analyses, TST never significantly predicted the sleep benefit in episodic memory.

SWS. We also investigated whether more time spent in SWS increases the sleep benefit in episodic memory as predicted by both consolidation and interference theories. This effect was not significant although descriptively in line with the expectation, $\beta = 0.05$, SE = 0.08, 95% CI [-0.13, 0.23], t(14.52) = 0.60, p = .558, k = 427, m = 123 (see Table S10). Informative hypothesis evaluation resulted in most support for H_0 (PMP = .88; BF_{ca} = BF_{ia} = 2.60, BF_{0a} = 17.36). Across all analyses performed, SWS never significantly predicted the sleep benefit in episodic memory.

REM Sleep. Consolidation as well as interference theories predict that longer REM sleep increases the sleep benefit. Again, however, the REM sleep effect was not statistically significant and was even opposite in direction, β = -0.016, SE = 0.12, 95% CI [-0.28, 0.25], t(15.63) = 0.13, p = .901, k = 427, m = 123 (see Table S11). Not surprisingly, H_0 received most support from the data in a subsequent informative hypothesis evaluation (PMP = .92; BF_{ca} = BF_{ia} = 0.81, BF_{0a} = 20.49). REM sleep never significantly predicted the sleep benefit in episodic memory in any of the analyses performed for this moderator.

Exploratory Variables

Stimulus Material. We investigated whether the sleep benefit in episodic memory generalizes across various stimulus materials. There were no statistically significant differences between the three moderator levels, F(3, 48.12) = 2.71, p = .076, k = 777, m = 254. Importantly, sleep in comparison to wakefulness improved episodic memory for verbal material, g = 0.46, t(94.31) = 0.46

10.94, p < .001, pictures, g = 0.44, t(39.37) = 7.98, p < .001, and mixed and other material, g = 0.33, t(63.65) = 6.70, p < .001 (see Figure 4J). This descriptive result pattern emerged in the majority of performed analyses. The exceptions showed a reverse order for verbal material and pictures in a few sensitivity analyses with a minimal advantage for pictures (see, for example, Tables S53 to S58).

Sleep Study Design. Next, we tested the influence of sleep study designs on the sleep benefit in episodic memory. As indicated by the AHT F test, there were statistically significant differences between the seven moderator levels, F(7, 27.80) = 2.77, p = .031, k = 777, m = 254. With the exception of total sleep deprivation with recovery night, g = 0.14, t(11.35) = 1.14, p = .278, all sleep study designs showed a significant sleep benefit in episodic memory. The largest sleep benefit occurred for nighttime naps, g = 0.54, t(10.18) = 2.61, p = .025, and natural sleep and wakefulness, g = 0.54, t(115.89) = 14.28, p < .001, followed by REM sleep deprivation, g = 0.43, t(11.87) = 6.45, p < .001, daytime naps, g = 0.34, t(37.64) = 6.01, p < .001, SWS deprivation, g = 0.31, t(12.04) = 2.33, p = .038, and finally total sleep deprivation without recovery night, g = 0.30, t(15.00) = 3.25, p = .005 (see Figure 4K). Although the result pattern varied across analyses, we constantly found descriptively larger sleep benefits in designs that did not impose constraints on SWS, such as nighttime naps, natural sleep and wakefulness, and REM sleep deprivation. To conclude, our findings urge caution in generalizing results across different sleep study designs.

Retention Interval. We also explored whether the length of the retention interval moderates the sleep benefit in episodic memory. In the main moderator analysis, this was clearly not the case, β = -0.001, SE = 0.002, 95% CI [-0.01, 0.01], t(2.11) = 0.37, p = .744, k = 777, m = 254. With the single exception of the "words and word pairs" subset in which retention interval length had a significantly negative effect on the sleep benefit (see Tables S8, S13, S18, S23, S28, S38, and S43), the length of the retention interval never significantly predicted the sleep benefit in episodic memory.

Experimental Design. As indicated by the AHT F test, between-subjects and within-subject designs did not differ significantly with respect to the sleep benefit observed, F(2, 76.54) = 1.87, p = .176, k = 777, m = 254. The sleep benefit was statistically significant in both between-subjects, g = .176

0.44, t(128.68) = 14.05, p < .001, and within-subject designs, g = 0.34, t(58.12) = 6.03, p < .001 (see Figure 4L). The same pattern was observed in all analyses performed for this moderator. Thus, there is no evidence that the sleep benefit in episodic memory is moderated by the experimental design.

Control Variables

We incorporated "randomization" and "accuracy of effect size computation" as control variables into the moderator analysis to help evaluate potential effects of uncertainties when coding primary studies on meta-analytic results. As is apparent in Table 2, neither "randomization" nor "accuracy of effect size computation" significantly moderated the sleep benefit in episodic memory. In none of the analyses performed we found noticeable effects of these control variables except for descriptive order reversals in some analyses (see, for example, Tables S8, S10, S11, and S58 to S60). We thus conclude that uncertainties in coding design aspects and extracting effect size estimates from primary studies did not affect our meta-analytic results.

Selective Reporting

As can be seen in Figure 5, the contour-enhanced funnel plot for the main data set did not show an obvious asymmetric pattern. This visual check was corroborated by the modified Egger's regression test. There was no statistically significant asymmetry for the main data set, $\beta = 0.534$, SE = 0.35, 95% CI [-0.15, 1.22], t(105.22) = 1.54, p = .126, k = 823, m = 271. However, previous research has shown that other variants of the modified Egger's regression test have only limited power to detect potential asymmetry in the funnel plot (Rodgers & Pustejovsky, 2020). Thus, even though there is no clear indication of funnel plot asymmetry, selective reporting may have affected our meta-analytic results. After correcting for potential funnel plot asymmetry using a modified variant of Egger's linear regression (Pustejovsky & Rodgers, 2019; Rodgers & Pustejovsky, 2020), the overall sleep benefit estimate was in fact only about half as large but still statistically significant in the main data set, g = 0.28, SE = 0.10, 95% CI [0.07, 0.48], t(103.71) = 2.69, p = .008, k = 823, m = 271.

In addition, we tested more directly for selective reporting by applying the 3PSM described in the Method section to subsets of our main data set that include independent effect sizes only. The three

parameters of the 3PSM and the likelihood ratio test are reported in Table 3 for each subset of the data. Across all 20 subsets analyzed, we found a median for heterogeneity between effect sizes of τ^2 = 0.11 (range: 0.02 – 0.47), a median for corrected overall effect size estimate of β = 0.41 (range: 0.05 – 0.79), and a median for the probability ratio of observing a published nonsignificant result relative to a significant one of ψ = 0.80 (range: 0.13 – 4.05). The 3PSM likelihood ratio test was statistically significant at the 5% level in three of the 20 subsets, slightly more often than expected when H_0 holds in each of the 20 tests. However, it must be taken into account that Type I error rates can be inflated in the 3PSM likelihood ratio test (Rodgers & Pustejovsky, 2020). Thus, the 3PSM result cannot be seen as clear evidence for selective reporting in the main data set.

In a nutshell, although the overall effect of sleep on episodic memory may have been overestimated in the present meta-analysis due to publication bias or other forms of selective reporting, we did not find strong evidence for selective reporting that would compromise the validity of our results. Importantly, a reliable sleep benefit persists even if potential selective reporting is accounted for.

Discussion

Despite a wealth of research on the sleep benefit in episodic memory that accumulated across almost one century, three important questions have not been fully clarified so far: How many different neurocognitive mechanisms contribute to the sleep benefit in episodic memory and what is their nature? How much does sleep boost episodic memory compared to wakefulness on average? And finally, which variables moderate the sleep benefit? Here, we provide the first comprehensive meta-analysis on the sleep benefit in episodic memory to address these questions.

First, we quantified the sleep benefit to inform about its practical significance and to prevent underpowered studies in future research. For this purpose, we fitted intercept-only three-level models with RVE to account for dependencies among effect sizes due to multiple outcome reporting and multiple group comparisons in primary studies. We found a moderate sleep benefit in episodic memory for the main data set (g = 0.44). Importantly, this overall effect size remained significant if

we corrected for potential selective reporting as reflected in funnel plot asymmetry but was somewhat reduced (g = 0.28). In addition, 3PSM analyses resulted in statistically significant corrected overall effect size estimates for all independent-data subsets except one. The median across all 20 subsets was 0.41 which deviates not much from the uncorrected effect size estimate of 0.44. Notably, our effect size estimate for the sleep benefit in episodic memory is quite similar to corresponding effect sizes for the sleep effect in prospective memory (d = 0.41, 95% CI [0.25, 0.56], k = 24; Leong et al., 2019) and motor memory (d = 0.44, 95% CI [0.09, 0.79], p = .018, k = 23, Pan & Rickard, 2015; g = 0.43, 95% CI [0.31, 0.55], p < .001, k = 53, Schmid et al., 2020). Also note that typical well-established memory effects documented in the literature are not much larger, and may be even smaller, for example the testing effect (g = 0.50, 95% CI [0.42, 0.58], p < .001, k = 159; Rowland, 2014), the retrieval-induced forgetting effect (g = 0.35, 95% CI [0.32, 0.38], p < .010, k = 512; Murayama et al., 2014), or the environmental context effect (g = 0.28, 95% CI [0.23, 0.33], k = 93; Smith & Vela, 2001). The sleep benefit in episodic memory can thus be considered a robust effect of similar size than other well-established effects in human memory research.

Second, we examined the impact of each of 18 moderator variables, statistically controlling for all others, by performing multiple-moderator analyses with RVE. For 12 of these moderator variables, hypotheses were previously derived from consolidation or interference theories. Six additional moderator candidates were more explorative in character. They were included to control for effects of study design features (4 variables) and uncertainties in coding primary studies (2 variables) on meta-analytic effect size estimates. These moderator analyses enabled important theoretical and methodological insights on the sleep benefit in episodic memory that will be discussed in detail below.

Third, we conducted a series of sensitivity analyses for different subsets of our data to evaluate how alternative methodological decisions might have impacted our results. By and large, the sensitivity analyses showed that all findings reported here are robust.

Taken together, the current meta-analysis informs about the relevance of the sleep benefit in episodic memory and provides valuable information about various moderators that can help researchers when designing studies and interpreting results. In the following section, we first discuss whether and how consolidation and interference theories can account for the results of the meta-analysis. On the basis of our findings, we then propose a theoretical framework that combines the strengths of both accounts. Next, methodological implications of our moderator analyses are summarized and limitations of the present meta-analysis are discussed. Finally, we suggest directions for future research.

Evaluating Consolidation and Interference Theories

Moderator variables of prime importance are those for which extant theories of the sleep benefit in episodic memory—consolidation and interference theories—make relatively clear-cut predictions. We first discuss specific predictions of the two theory classes before we move on to predictions shared by both accounts. The main question of interest in our discussion is whether results are descriptively in line with the predictions, and whether the predicted pattern is observed constantly across all subgroup and sensitivity analyses performed. Of course, statistical significance of moderator effects is of relevance also. However, given power limitations due to small numbers of effect sizes observed for some of the moderator levels that entered into our analyses—a limitation we unfortunately could not avoid given the current state of research on the sleep benefit in episodic memory (see Limitations section)—, statistical significance should not be treated as the main criterion of interest. We aim at evaluating the evidence for and against each theoretical framework as broadly as possible, even if some of the conclusions should be considered preliminary and thus require experimental follow-up evaluations for establishing their truth status more rigorously.

Specific Predictions of Consolidation Theories

Consolidation theories can explain observed moderator effects of emotionality and directed forgetting instructions quite well. Like previous meta-analyses (Lipinska et al., 2019; Schäfer et al., 2020), we failed to find a statistically significant moderator effect of emotionality. However, we

observed a descriptive pattern in the direction predicted by consolidation theories. Because this support for consolidation theories is only tentative, more research on negative and positive memories is needed. In addition, the valence and arousal dimensions of emotional stimuli are likely confounded in this meta-analysis, because primary studies did not control for arousal and provided only valence but no arousal data. Future research should therefore disentangle the two dimensions of emotionality more rigorously than past research has done.

Similar to emotionality, effects of directed forgetting instructions were descriptively in line with consolidation theories but not statistically significant. However, little is known about (a) the mechanisms by which the hypothesized synaptic tags increase memory consolidation during sleep, and (b) the cognitive processes contributing to the directed forgetting effect in different paradigms. These questions need to be scrutinized in further research. Furthermore, future research should test more directly whether and how directed forgetting instructions affect storage and retrieval processes involved in the sleep benefit in episodic memory.

Specific Predictions of Interference Theories

Interference theories can account for the observed meta-analytic results concerning effects of restudy, testing, feedback, and retrieval procedure quite nicely. For restudy, testing, and feedback, however, further research is needed to uncover the underlying processes and to examine the possible additional impact of restudy and testing procedures (e.g., number of restudy or testing cycles, application of learning criteria).

Although the results for encoding instruction are not in line with the expectation, interference theories offer a plausible post-hoc explanation for the observed result pattern. Specifically, memory strength following intentional learning might be so pronounced in many studies that most items surpass not only the lower retrieval threshold after sleep but also the higher threshold after wakefulness. This reduces the sleep benefit in episodic memory retrieval for intentional compared to incidental learning. The assumption that sleep benefits episodic memory retrieval less the better stimuli are learned is supported by studies showing that retroactive interference decreases with

higher study efforts (Howe, 1970; McGeoch, 1929). However, additional studies are necessary to test the proposed explanation for effects of intentional versus incidental learning on the sleep benefit via memory retrieval processes more thoroughly.

Interference theories also explain why the sleep benefit becomes more apparent the more difficult memory retrieval is in a memory test. In line with interference theories, we found the largest sleep benefit in free recall, followed by cued recall, and lastly recognition tasks. Whereas this result pattern was not statistically significant in the main data set, it was more pronounced and statistically significant in the subset of studies employing verbal learning materials (i.e., single words and word pairs). This difference in results can be explained by the relative strong imbalance of different retrieval procedures in the main data set. Specifically, compared to cued recall and recognition, the moderator level "free recall" contains few effect sizes in the main data set. This introduces uncertainty in effects size estimates and may thus lead to nonsignificant differences between moderator levels. In contrast, effect sizes in the subset of single words and word pairs are more uniformly distributed across the three moderator levels. Importantly, except for the main sensitivity analysis on memory performance measures (see Table S17 in the supplemental materials), all analyses showed the descriptively largest sleep benefit in free recall, followed by cued recall, and lastly recognition.

Even though we cannot completely rule out that possible side-effects of systems consolidation are involved in higher sleep benefits for recall tests (see the paragraph on the focal variable "retrieval procedure" in the Potential Moderators section), interference theories offer a plausible explanation of why the sleep benefit is moderated by the memory test used. According to interference theories, sleep improves memory retrieval by reducing retroactive interference.

Because free recall relies more heavily on memory retrieval than cued recall, and cued recall more than recognition, the sleep benefit should be largest in free recall, followed by cued recall, and lastly recognition tasks (cf. Dyne et al., 1990; McKinney, 1935; Postman, 1952).

By suggesting that retrieval procedure moderates the sleep benefit for single words and word pairs, our findings are in line with a recent meta-analysis that focused on the Deese-Roediger-McDermott (DRM) paradigm (Roediger & McDermott, 1995). Newbury and Monaghan (2019) observed a significant moderator effect of recall versus recognition for correctly remembered words; Q(1) = 3.93, p = .047, k = 13. However, in contrast to our results, no significant sleep benefit was found, neither in recall (g = 0.41, 95% CI [-0.09, 0.91], p = .112, k = 5) nor in recognition tasks (g = 0.01, 95% CI [-0.19, 0.20], p = .958, k = 8; Newbury & Monaghan, 2019), probably because of the special nature of the DRM paradigm. This notwithstanding, the retrieval procedure appears to affect the sleep benefit in episodic memory not only in our study. Given this evidence, it seems reasonable to use recognition tests primarily to assess sleep benefits in episodic memory storage and (free) recall tests to assess sleep benefits in episodic memory retrieval. Because recognition tasks are least reliant on memory retrieval in general, they might be best suited to capture sleep benefits in storage without associated benefits in retrieval. In contrast, (free) recall test heavily rely on retrieval processes and are thus best suited to measure sleep benefits in memory retrieval.

Predictions Shared by Consolidation and Interference Theories

Both consolidation and interference theories are roughly in line with the results for age and also consistent with the findings of a previous meta-analysis (Gui et al., 2017). However, to reliably decide whether the sleep benefit in episodic memory monotonically decreases across the entire life span, more studies on children, adolescents, middle aged adults, and older adults are required. In addition, provided that sufficient data is available, future meta-analyses should consider age as a continuous variable to capture the course of the sleep benefit across the life-span in a more fine-graded fashion.

Finally, both consolidation and interference theories would predict that increases in TST, SWS, and REM sleep are accompanied by increases in the sleep benefit in episodic memory. However, we observed no significant effects on the sleep benefit whatsoever, neither for TST, nor for SWS, nor for REM sleep duration. As such, our results for SWS are in line with a recent study showing no

statistically significant relationship between sleep benefits and the amount of SWS (Cordi & Rasch, 2021b). In the current meta-analysis, we cannot rule out the possibility that interaction effects between sleep parameters, sleep study designs, and emotionality are responsible for the null effects observed. Specifically, sleep study designs differ in terms of TST, amount of SWS, and REM sleep duration. In fact, some of the sleep study designs are systematically tailored to study effects of specific sleep stages, resulting in a confound between sleep parameters and sleep study designs. In addition, REM sleep has been assumed to foster storage of emotional and negative memories in particular, resulting in a possible additional confound between REM sleep and effects of emotionality or negativity of memories. In a meta-analysis, such complex interactions are best investigated by examining the effect of sleep parameters in different subsets for each sleep study design or for each combination of sleep study designs and emotionality. However, this was impossible for our current meta-analysis because the data set did not provide a sufficiently large number of effect sizes. Hence, this open problem must be left for future research.

In sum, although several aspects of our moderator analyses call for further research, we found encouraging support for predictions of both classes of theories. Some results are more in accordance with consolidation theories whereas others more closely match predictions of interference theories. This suggests that both theoretical accounts should not be considered as mutually exclusive but rather be treated in an integrative fashion. In the following section, we propose such an integrative account that combines core principles of consolidation and interference theories.

An Integrative Framework That Unifies Consolidation and Interference Theories

Whereas consolidation theories have been very influential in neuroscience, they have been less so in psychology with respect to memory and sleep research (cf. Wixted, 2004). The opposite holds for interference theories: Although these theories have always been prominent in psychology, they have rarely been considered in neuroscientific research on the sleep benefit in episodic memory. Rather, results have typically been interpreted in terms of active sleep effects on memory storage only, thereby neglecting a possible passive effect of sleep on memory retrieval altogether.

One possible reason for this imbalance may be a diverging understanding of the notion of interference. As outlined above, psychological interference theories propose that sleep passively improves memory retrieval by protecting memories from retroactive interference. In other words, sleep supports the accessibility of successfully stored episodic memories. In neuroscience, however, a possible passive effect of sleep on episodic memory is almost exclusively discussed in terms of memory storage. Specifically, it has been assumed that sleep in comparison to wakefulness either creates an ideal condition for memory consolidation (opportunistic consolidation hypothesis) or that it temporarily preserves the memory representation from decay by reducing retroactive interference (Ellenbogen, Payne, & Stickgold, 2006; Mednick et al., 2011). To our knowledge, a positive influence of sleep on retrieval processes in addition to active (or passive) effects of sleep on memory storage has hardly been ever discussed by neuroscientists.

On the basis of the present meta-analytic results, we argue that it is important to integrate consolidation theories, the opportunistic consolidation hypothesis, and interference theories within a broader theoretical framework to improve our understanding of the joint contribution of all processes underlying the sleep benefit in episodic memory. Limiting the theoretical debate to eitheror questions such as "Does sleep improve memory by fostering consolidation or by reducing interference?" or "Are sleep effects on memory due to active or passive processes?" is likely to be misleading. Sleep can indeed be beneficial for all these reasons. That is, sleep can foster consolidation and reduce interference simultaneously, thereby improving episodic memories in active and passive ways at the same time. There is no theoretical or empirical reason why consolidation and interference theories should be treated as mutually exclusive, as has been done so far in research on sleep-induced memory benefits.

We thus propose to abandon the either-or debate when trying to explain sleep effects on memory and to replace it by a debate that draws on an integrative, multicausal framework which includes key assumptions of both consolidation and interference theories. In essence, such an integrative framework is based on two core assumptions:

- Sleep compared to wakefulness actively strengthens episodic memory storage due to memory consolidation, perhaps additionally supported by passive protection from interference.
- In parallel, sleep also passively supports episodic memory retrieval by reducing retroactive interference during sleep.

In support of our integrative framework, we found a significant sleep benefit across all memory tests investigated, with largest sleep benefits in free recall, followed by cued recall, and lastly recognition when memory for word materials is tested. This simple result pattern can hardly be explained by consolidation or interference theories considered in isolation, but perfectly by a combination of both. Specifically, according to consolidation theories, sleep compared to wakefulness should improve episodic memory storage irrespective of the retrieval procedure in the memory test that follows. This explains why the sleep benefit is significant across all types of memory tests used. In addition, according to interference theories, sleep compared to wakefulness should improve episodic memory retrieval by reducing retroactive interference. Because memory tests differ in how much they rely on memory retrieval, the size of the sleep benefit in episodic memory retrieval should depend on the type of test used. This explains why the sleep benefit is largest in free recall, followed by cued recall, and lastly recognition tests.

Furthermore, our results for testing and encoding instructions also point into the same direction.

To reiterate, the sleep benefit did neither differ significantly between testing and no testing conditions nor between intentional and incidental learning conditions. Because the sleep benefit was significant not only for both levels of testing and encoding instructions but also for both levels of restudy and feedback, it appears that sleep compared to wakefulness improves memory storage to similar degrees, irrespective of encoding conditions. In addition, interference theories offer a plausible explanation why the sleep benefit is not moderated by testing but by restudy and feedback. Moreover, interference theories also offer an explanation for a smaller sleep benefit in intentional- than incidental-learning conditions. Assuming that intentional learning increases

memory strength more than incidental learning, the memory strength after intentional learning might in fact become so strong that most items surpass not only the lower retrieval threshold after sleep but also the higher threshold after wakefulness. This in turn may reduce the sleep benefit for intentional compared to incidental learning.

To conclude, we propose an integrative framework by combining theories that originate from neuroscience and psychology. In a nutshell, our integrative framework posits that sleep compared to wakefulness affects episodic memory in two qualitatively different ways: It actively increases memory storage due to memory consolidation and it also passively improves memory retrieval by reducing interference. The results of the current meta-analysis support our integrative framework by showing that an exhaustive explanation of the full spectrum of results requires consideration of both consolidation and interference theories. Critically, our findings also suggest that memory consolidation strengthens storage of episodic memories equally, irrespective of their encoding strength. Although single theoretical accounts can offer plausible alternative explanations that are important to consider (cf. Diekelmann et al., 2009), the integrative framework currently provides the most straightforward explanation for the complex pattern of results. Future research is required to rigorously test the predictions of the integrative framework while accounting for possible alternative explanations.

Methodological Implications

We investigated whether the definition of the dependent variable impacts the sleep benefit in episodic memory. In line with our expectation, we observed a larger sleep benefit for direct measures of retention (i.e., differences between immediate and delayed test performance) than for measures of delayed test memory performance. Such difference measures of retention control for individual differences in learning abilities by default and are thus more reliable than measures based on delayed tests only. Thus, if the research question and the study design permits, researchers should prefer direct measures of retention to increase the standardized effect size of the sleep benefit and thus the power of the statistical test.

Moreover, we observed that the sleep benefit differs considerably between the various ways to compare sleep versus wakefulness (see Figure 4K). Notably, across performed subgroup and sensitivity analyses, nighttime naps, natural sleep and wakefulness, and REM sleep deprivation²⁰ constantly showed the largest sleep benefits compared to other sleep study designs. This pattern fits well to the assumption that SWS is important for the sleep benefit. However, it might not be SWS duration per se, but rather specific SWS characteristics such as sleep spindles that matter most (Cordi et al., 2021a). In any case, researchers should avoid unsubstantiated generalization of findings across various sleep study designs.

We additionally explored whether the sleep benefit is moderated by stimulus materials and experimental design (i.e., between-subjects vs. within-subject design) but found no convincing evidence for such effects. To conclude, the sleep benefit in episodic memory appears to be moderated by methodological decisions regarding memory assessment (i.e., definition of dependent variable) and study design (i.e., sleep study design) but less so by other methodological aspects.

Limitations

Although, according to our knowledge, the present meta-analysis on sleep and memory appears to be the largest in terms of the number of studies, effect sizes, and moderators included, not every moderator variable of potential interest has been addressed. For some additional moderators, the available sample of effect sizes was simply too small. Others exhibited too many missing values to be included in a multiple-moderator analysis. For example, it would be informative to investigate motivational aspects such as reward and the expectation that a memory is relevant for the future because consolidation theories predict that the sleep benefit is affected by such aspects of the encoding situation (Diekelmann et al., 2009; Stickgold & Walker, 2013). Because we could not address these questions on the basis of the available effect sizes in our data set, future research is needed to fill this gap.

²⁰ To reiterate, in REM sleep deprivation studies participants either sleep in the first part of the night to reduce REM sleep or are awakened every time the first epoch of REM sleep occurs throughout a full night of sleep.

Closely related, for the analysis of several moderators and moderator levels, the available number of effect sizes was relatively small, introducing uncertainty in the corresponding analysis. In particular, nonsignificant effects in our moderator analyses should be considered with caution. For subsequent research, we recommend experimental manipulations of moderators of interest, using an overall sample size that ensures high power. Consider, for example, the prediction of the opportunistic consolidation hypothesis that the sleep benefit in episodic memory is diminished by interference-reducing states like quiet wake. To test this prediction, we performed an exploratory moderator analysis (see Footnote 2 on p. 16) but were not able to draw solid conclusions because the number of available effect sizes for wakeful rest was simply too small for a meaningful interpretation. For future research, we therefore recommend a statistically more powerful direct comparison of the three experimental conditions "sleep", "low interference wake state", and "high interference wake state" while keeping the length of the retention interval constant.

Another limitation of the current meta-analysis is that it includes few unpublished data. As such, the inclusion of unpublished data is desirable because it prevents effect sizes from being systematically overestimated as a consequence of selective reporting. However, we statistically checked whether there is evidence for selective reporting in the data set of the present meta-analysis. We used a modified variant of Egger's linear regression to test for funnel plot asymmetry and applied the 3PSM to check for selective reporting more directly. These analyses revealed no significant funnel plot asymmetry and no clear evidence for publication bias. Nevertheless, some selective reporting might have been involved in the studies that entered our meta-analysis. However, even if we control for this possibility, a significant sleep benefit of moderate size remained for our main data set.

Beyond the limited amount of gray literature included, the validity and generalizability of the meta-analytic results may be affected by the selection criteria we employed (cf. Johnson, 2021). First, we included only articles written in English or German. Based on this selection criterion, 31 articles were excluded (see Figure 2). Note, however, that all these articles would have been

excluded for reasons other than language as well (e.g., no primary empirical results reported, no accessible full text). We therefore do not believe that our conclusions are compromised by a systematic language bias (see the OSF for a complete list of articles written in other languages and the respective reasons for exclusion;

https://osf.io/dr6hw/?view_only=c55b3737b6c642cfb170276088930a04).

Second, except for Pawlizki's (2012) dissertation, all articles included are in English, suggesting an English-centered sample. However, this interpretation is not in line with the fact that almost one third of the articles originate from Germany. This notwithstanding, although we included samples from 23 countries all over the world, most samples are from western or western-influenced cultures, which could introduce sampling bias. Importantly, however, this potential bias does not result from our search strategy or study selection criteria but is a necessary consequence of the available literature. As such, the sampling bias in psychology and neuroscience towards western, educated, industrialized, rich, and democratic (WEIRD) populations is well known (e.g., Falk et al., 2013; Henrich et al., 2010). Thus, as our included articles are not and cannot be representative for the world population, more research in non-WEIRD populations and direct comparisons between WEIRD and non-WEIRD populations is desirable to check whether the conclusions drawn from this meta-analysis generalize to non-western populations.

Finally, we focused on healthy human individuals using behavioral measures. Consequently, our meta-analytic results are silent about clinical populations or physiological data from neuroimaging and polysomnography. However, considering clinical populations and including physiological data may contribute to a better understanding of the neurocognitive mechanisms underlying the sleep benefit. Therefore, future meta-analyses should consider the additional inclusion of clinical populations and physiological data. To help overcoming the limitations outlined above, we provide our complete data set and codebook

(https://osf.io/dr6hw/?view_only=c55b3737b6c642cfb170276088930a04) so that it can serve as a starting point for future evidence syntheses.

Directions for Future Research

Throughout the Discussion section, we have already suggested various directions for future research. Here we focus on two remaining issues. First, we elaborate on how future research can use the results of the current meta-analysis to increase the replicability of the sleep benefit in episodic memory. Second, on the basis of our results, we address how the underlying processes of the sleep benefit can be further investigated in the future.

Replicability of results is commonly considered to be one of the cornerstones of science (Popper, 2002). In behavioral research, replication success is typically defined in terms of rejecting the null hypothesis using statistical tests. Hence, a successful replication depends on statistical power, heterogeneity between studies, and possible selective reporting, among other factors (Stanley et al., 2018). If the statistical power is low in either the original or the replication study, chances of successful replication are slim. Due to time and cost intensive methods in sleep research (e.g., overnight EEG studies), sample sizes are typically quite small. For example, when considering the studies included in the present meta-analysis, the median number of participants is approximately 16 per experimental condition. Such small sample sizes make it quite unlikely to detect an effect that is actually present. To illustrate, with an error probability α of .05, the probability to detect a mean difference of medium size (i.e., d = 0.50) with a two-tailed t test comparing two independent groups, each containing 16 participants (N = 32) is only 28% (= statistical power, 1 - β). Thus, low statistical power is one of the reasons why replication attempts often fail. To solve this problem, it is necessary to conduct studies with sufficiently large sample sizes. Methods such as a priori power analysis (e.g., Faul et al., 2007) and sequential hypotheses testing (e.g., Lakens, 2014; Lakens & Evers, 2014; Schnuerch & Erdfelder, 2020; Schönbrodt et al., 2017) ensure adequate power of statistical hypothesis tests by controlling both Type I (α) and Type II (β) error rates. These methods are most informative when researchers have a convincing rationale for the effect size of the phenomenon under investigation. Our meta-analysis informs about these effect sizes and thus helps to increase

the statistical power and consequently the replicability in the future, provided that sleep researchers plan their test strategies and sample sizes accordingly.

Another factor that threatens replication success is heterogeneity between studies. Sleep researchers use a wide variety of populations, materials, sleep study designs, and measures. For example, the effect has been investigated with different stimuli (e.g., words, pictures) and sleep study designs (e.g., naps, total sleep deprivation). Whereas this heterogeneity between studies is desirable to demonstrate generalizability of the effect across different settings, it is problematic for replications. If crucial methodological characteristics differ between original studies and direct replication attempts, chances to successfully replicate previous findings diminish. In other words, methodological characteristics can act as moderators that increase or decrease the size of the effect. Knowledge about effects of moderator variables can help researchers to design studies and interpret unexpected or (apparently) contradictory results appropriately. Because the current meta-analysis provides important information about moderators and enables researchers to equate replication and original studies with respect to these moderators, it facilitates successful replications in future studies.

Although our meta-analysis provides valuable information about the sleep benefit in episodic memory, the validity of its results might be threatened by selective reporting. Selective reporting caused by publication bias, among others, restricts the complete set of primary studies to a subset that includes significant results in the hypothesized direction only. Having access only to this subset necessarily leads to overestimation of effect size. Hence, selective reporting not only produces distorted impressions of an effect, but also reduces the probability of a successful replication. For this reason, we checked for selective reporting and estimated a corrected overall effect size for the sleep benefit in episodic memory. However, the best remedy for distortions of the empirical evidence by selective reporting is to eliminate its causes. Preregistration will likely be an important tool in future research on sleep-induced memory benefits.

Most importantly, the integrative framework of consolidation and interference theories proposed here offers new impulses for future research and theory building. According to this account, sleep fosters memory in two ways, namely, by improving memory storage due to memory consolidation and by improving memory retrieval by reducing interference. How can memory storage and retrieval be disentangled, appropriately measured, and investigated more thoroughly? One way to proceed would be to use appropriately designed multinomial processing tree (MPT) models that aim at uncontaminated measures of storage and retrieval processes underlying successful and unsuccessful memory performance (Batchelder & Riefer, 1999; Erdfelder et al., 2009).

Conclusion

Almost a century of research indicates that sleep benefits episodic memory. The current metaanalysis confirms this well-established finding and shows that the sleep benefit is of moderate size.

To advance our understanding of the processes contributing to the sleep benefit in episodic
memory, we integrated prominent theories from different research areas. In our meta-analysis, we
found encouraging initial support that sleep benefits storage in memory due to memory
consolidation and, in addition, retrieval from memory by reducing interference. Beyond these
theoretical contributions, the present meta-analysis offers practical guidance to increase the
replicability of the sleep benefit in future research. We identified crucial methodological
characteristics that, if not adequately considered, could jeopardize replication success. More
specifically, we showed that the sleep benefit in episodic memory is moderated by the memory
measure used and the sleep study design. In sum, although further research is needed, our metaanalysis provides convincing evidence that sleep compared to wakefulness following learning
improves both storage in and retrieval from episodic memory significantly. Clearly, sleep not only
benefits health and well-being in everyday life, but is also essential for proper episodic memory
functioning.

References

- References marked with an asterisk indicate articles included in the meta-analysis.
- *Abel, M. (2011). Sleep-associated and retrieval-associated memory [Doctoral dissertation, University of Regensburg]. University Library of Regensburg. https://epub.uni-regensburg.de/24260/
- *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., & Bäuml, K.-H. T. (2019). List-method directed forgetting after prolonged retention interval: Further challenges to contemporary accounts. *Journal of Memory and Language*, *106*, 18–28. https://doi.org/10.1016/j.jml.2019.02.002
- *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
- *Alger, S. E., Lau, H., & Fishbein, W. (2010). Delayed onset of a daytime nap facilitates retention of declarative memory. *PLOS ONE*, *5*(8), Article e12131. https://doi.org/10.1371/journal.pone.0012131

- *Alger, S. E., Lau, H., & Fishbein, W. (2012). Slow wave sleep during a daytime nap is necessary for protection from subsequent interference and long-term retention. *Neurobiology of Learning and Memory*, *98*(2), 188–196. https://doi.org/10.1016/j.nlm.2012.06.003
- *Alger, S. E., & Payne, J. D. (2016). The differential effects of emotional salience on direct associative and relational memory during a nap. *Cognitive, Affective, & Behavioral Neuroscience*, *16*, 1150–1163. https://doi.org/10.3758/s13415-016-0460-1
- *Aly, M., & Moscovitch, M. (2010). The effects of sleep on episodic memory in older and younger adults. *Memory*, *18*(3), 327–334. https://doi.org/10.1080/09658211003601548
- 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
- American Psychological Association. (n.d.). *Thesaurus of Psychological Index Terms*. Retrieved January 9, 2018, from https://www.apa.org/pubs/databases/training/thesaurus
- 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
- 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., Jefferies, E., & Gaskell, M. G. (2018). A role for consolidation in cross-modal category learning. *Neuropsychologia*, *108*, 50–60. https://doi.org/10.1016/j.neuropsychologia.2017.11.010
- *Ashworth, A., Hill, C. M., Karmiloff-Smith, A., & Dimitriou, D. (2014). Sleep enhances memory consolidation in children. *Journal of Sleep Research*, *23*(3), 304–310. https://doi.org/10.1111/jsr.12119
- *Ashworth, A., Hill, C. M., Karmiloff-Smith, A., & Dimitriou, D. (2017). A cross-syndrome study of the differential effects of sleep on declarative memory consolidation in children with

- neurodevelopmental disorders. *Developmental Science*, *20*(2), Article e12383. https://doi.org/10.1111/desc.12383
- Astill, R. G., Van der Heijden, K. B., Van IJzendoorn, M. H., & Van Someren, E. J. W. (2012). Sleep, cognition, and behavioral problems in school-age children: A century of research meta-analyzed. *Psychological Bulletin*, *138*(6), 1109–1138. https://doi.org/10.1037/a0028204
- *Atherton, K. E., Nobre, A. C., Zeman, A. Z., & Butler, C. R. (2014). Sleep-dependent memory consolidation and accelerated forgetting. *Cortex*, *54*, 92–105. https://doi.org/10.1016/j.cortex.2014.02.009
- Atherton, L. A., Dupret, D., & Mellor, J. R. (2015). Memory trace replay: The shaping of memory consolidation by neuromodulation. *Trends in Neurosciences*, *38*(9), 560–570. https://doi.org/10.1016/j.tins.2015.07.004
- *Atienza, M., & Cantero, J. L. (2008). Modulatory effects of emotion and sleep on recollection and familiarity. *Journal of Sleep Research*, *17*(3), 285–294. https://doi.org/10.1111/j.1365-2869.2008.00661.x
- *Backhaus, J., Hoeckesfeld, R., Born, J., Hohagen, F., & Junghanns, K. (2008). Immediate as well as delayed post learning sleep but not wakefulness enhances declarative memory consolidation in children. *Neurobiology of Learning and Memory*, *89*(1), 76–80. https://doi.org/10.1016/j.nlm.2007.08.010
- *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

- Bäuml, K.-H., Pastötter, B., & Hanslmayr, S. (2010). Binding and inhibition in episodic memory—

 Cognitive, emotional, and neural processes. *Neuroscience & Biobehavioral Reviews*, *34*(7), 1047—

 1054. https://doi.org/10.1016/j.neubiorev.2009.04.005
- *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
- *Baran, B., Mantua, J., & Spencer, R. M. C. (2016). Age-related changes in the sleep-dependent reorganization of declarative memories. *Journal of Cognitive Neuroscience*, *28*(6), 792–802. https://doi.org/10.1162/jocn_a_00938
- *Baran, B., Pace-Schott, E. F., Ericson, C., & Spencer, R. M. C. (2012). Processing of emotional reactivity and emotional memory over sleep. *Journal of Neuroscience*, *32*(3), 1035–1042. https://doi.org/10.1523/JNEUROSCI.2532-11.2012
- *Baran, B., Wilson, J., & Spencer, R. M. C. (2010). REM-dependent repair of competitive memory suppression. *Experimental Brain Research*, 203, 471–477. https://doi.org/10.1007/s00221-010-2242-2
- 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
- Barham, M. P., Enticott, P. G., Conduit, R., & Lum, J. A. G. (2016). Transcranial electrical stimulation during sleep enhances declarative (but not procedural) memory consolidation: Evidence from a meta-analysis. *Neuroscience & Biobehavioral Reviews*, *63*, 65–77. https://doi.org/10.1016/j.neubiorev.2016.01.009
- Barry, D. N., Coogan, A. N., & Commins, S. (2016). The time course of systems consolidation of spatial memory from recent to remote retention: A comparison of the immediate early genes Zif268, c-Fos and Arc. *Neurobiology of Learning and Memory, 128,* 46–55. https://doi.org/10.1016/j.nlm.2015.12.010

- 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
- Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, *41*(2), 257–278. https://doi.org/10.1111/j.2044-8317.1988.tb00901.x
- *Bell, M. C., Kawadri, N., Simone, P. M., & Wiseheart, M. (2014). Long-term memory, sleep, and the spacing effect. *Memory*, *22*(3), 276–283. https://doi.org/10.1080/09658211.2013.778294
- *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
- *Bennion, K. A., Payne, J. D., & Kensinger, E. A. (2017). Residual effects of emotion are reflected in enhanced visual activity after sleep. *Cognitive, Affective, & Behavioral Neuroscience*, *17*, 290–304. https://doi.org/10.3758/s13415-016-0479-3
- *Bennion, K. A., Steinmetz, K. R. M., Kensinger, E. A., & Payne, J. D. (2015). Sleep and cortisol interact to support memory consolidation. *Cerebral Cortex*, *25*(3), 646–657. https://doi.org/10.1093/cercor/bht255
- *Benson, K., & Feinberg, I. (1975). Sleep and memory: Retention 8 and 24 hours after initial learning.

 *Psychophysiology, 12(2), 192–195. https://doi.org/10.1111/j.1469-8986.1975.tb01275.x
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons. https://doi.org/10.1002/9780470743386
- 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
- Boyce, R., Williams, S., & Adamantidis, A. (2017). REM sleep and memory. *Current Opinion in Neurobiology*, 44, 167–177. https://doi.org/10.1016/j.conb.2017.05.001

- Brodt, S., Gais, S., Beck, J., Erb, M., Scheffler, K., & Schönauer, M. (2018). Fast track to the neocortex:

 A memory engram in the posterior parietal cortex. *Science*, *362*(6418), 1045–1048.

 https://doi.org/10.1126/science.aau2528
- Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, 114(3), 539–576. https://doi.org/10.1037/0033-295X.114.3.539
- *Brown, R. M., & Robertson, E. M. (2007). Off-line processing: Reciprocal interactions between declarative and procedural memories. *Journal of Neuroscience*, *27*(39), 10468–10475. https://doi.org/10.1523/JNEUROSCI.2799-07.2007
- Butler, A. C., Karpicke, J. D., & Roediger, H. L., III (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology:*Learning, Memory, and Cognition, 34(4), 918–928. https://doi.org/10.1037/0278-7393.34.4.918
- 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., Durrant, S. J., Musgrove, H., & Lewis, P. A. (2011). Sleep and environmental context:

 Interactive effects for memory. *Experimental Brain Research*, *214*, 83–92.

 https://doi.org/10.1007/s00221-011-2808-7
- *Cairney, S. A., Lindsay, S., Paller, K. A., & Gaskell, M. G. (2018). Sleep preserves original and distorted memory traces. *Cortex*, *99*, 39–44. https://doi.org/10.1016/j.cortex.2017.10.005
- 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
- *Campanella, C. (2014). The effect of sleep on memory consolidation of emotional and distinctive composite (Publication No. 3688108) [Doctoral dissertation, Emory University]. ProQuest Dissertations and Theses Global.

- Carr, M. F., Jadhav, S. P., & Frank, L. M. (2011). Hippocampal replay in the awake state: A potential substrate for memory consolidation and retrieval. *Nature Neuroscience*, *14*(2), 147–153. https://doi.org/10.1038/nn.2732
- Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology:

 A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological*Science, 2(2), 115–144. https://doi.org/10.1177/2515245919847196
- *Cellini, N., Torre, J., Stegagno, L., & Sarlo, M. (2016). Sleep before and after learning promotes the consolidation of both neutral and emotional information regardless of REM presence.

 Neurobiology of Learning and Memory, 133, 136–144. https://doi.org/10.1016/j.nlm.2016.06.015
- *Chambers, A. M., & Payne, J. D. (2014). Laugh yourself to sleep: Memory consolidation for humorous information. *Experimental Brain Research*, *232*, 1415–1427. https://doi.org/10.1007/s00221-013-3779-7
- *Chan, S., Pressler, R., Boyd, S. G., Baldeweg, T., & Cross, J. H. (2017). Does sleep benefit memory consolidation in children with focal epilepsy? *Epilepsia*, *58*(3), 456–466. https://doi.org/10.1111/epi.13668
- Chatburn, A., Lushington, K., & Kohler, M. J. (2014). Complex associative memory processing and sleep: A systematic review and meta-analysis of behavioural evidence and underlying EEG mechanisms. *Neuroscience & Biobehavioral Reviews*, *47*, 646–655. https://doi.org/10.1016/j.neubiorev.2014.10.018
- *Cherdieu, M., Reynaud, E., Uhlrich, J., Versace, R., & Mazza, S. (2014). Does age worsen sleep-dependent memory consolidation? *Journal of Sleep Research*, *23*(1), 53–60. https://doi.org/10.1111/jsr.12100
- Cirelli, C., & Tononi, G. (2015). Sleep and synaptic homeostasis. *Sleep*, *38*(1), 161–162. https://doi.org/10.5665/sleep.4348
- Coburn, K. M., & Vevea, J. L. (2015). Publication bias as a function of study characteristics. *Psychological Methods*, 20(3), 310–330. https://doi.org/10.1037/met0000046

- Coburn, K. M., & Vevea, J. L. (2019). weightr: Estimating weight-function models for publication bias (Version 2.0.2) [R package]. https://cran.r-project.org/package=weightr
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587
- *Coolidge, F. L. (1974). *Memory consolidation as a function of sleep and the circadian rhythm*(Publication No. 7516368) [Doctoral dissertation, University of Florida]. ProQuest Dissertations and Theses Global.
- Cordi, M. J., & Rasch, B. (2021a). How robust are sleep-mediated memory benefits? *Current Opinion in Neurobiology, 67*, 1–7. https://doi.org/10.1016/j.conb.2020.06.002
- Cordi, M. J., & Rasch, B. (2021b). No evidence for intra-individual correlations between sleep-mediated declarative memory consolidation and slow-wave sleep. *Sleep*, Article zsab034.

 Advance online publication. https://doi.org/10.1093/sleep/zsab034
- *Cox, R., Tijdens, R. R., Meeter, M. M., Sweegers, C. C. G., & Talamini, L. M. (2014). Time, not sleep, unbinds contexts from item memory. *PLOS ONE*, *9*(2), Article e88307. https://doi.org/10.1371/journal.pone.0088307
- 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
- *Cunningham, T. J., Crowell, C. R., Alger, S. E., Kensinger, E. A., Villano, M. A., Mattingly, S. M., & Payne, J. D. (2014). Psychophysiological arousal at encoding leads to reduced reactivity but enhanced emotional memory following sleep. *Neurobiology of Learning and Memory, 114*, 155–164. https://doi.org/10.1016/j.nlm.2014.06.002
- *Darsaud, A., Dehon, H., Lahl, O., Sterpenich, V., Boly, M., Dang-Vu, T., Desseilles, M., Gais, S., Matarazzo, L., Peters, F., Schabus, M., Schmidt, C., Tinguely, G., Vandewalle, G., Luxen, A., Maquet, P., & Collette, F. (2011). Does sleep promote false memories? *Journal of Cognitive Neuroscience*, 23(1), 26–40. https://doi.org/10.1162/jocn.2010.21448

- *Deliens, G., Gilson, M., Schmitz, R., & Peigneux, P. (2013). Sleep unbinds memories from their emotional context. *Cortex*, 49(8), 2221–2228. https://doi.org/10.1016/j.cortex.2012.11.014
- *Deliens, G., Leproult, R., Neu, D., & Peigneux, P. (2013). Rapid eye movement and non-rapid eye movement sleep contributions in memory consolidation and resistance to retroactive interference for verbal material. *Sleep*, *36*(12), 1875–1883. https://doi.org/10.5665/sleep.3220
- *Deliens, G., & Peigneux, P. (2014). One night of sleep is insufficient to achieve sleep-to-forget emotional decontextualisation processes. *Cognition and Emotion*, *28*(4), 698–706. https://doi.org/10.1080/02699931.2013.844105
- *Deliens, G., Schmitz, R., Caudron, I., Mary, A., Leproult, R., & Peigneux, P. (2013). Does recall after sleep-dependent memory consolidation reinstate sensitivity to retroactive interference? *PLOS ONE*, *8*(7), Article e68727. https://doi.org/10.1371/journal.pone.0068727
- 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.
 - http://www.learnmem.org/cgi/doi/10.1101/lm.051383.120
- *Diekelmann, S., Biggel, S., Rasch, B., & Born, J. (2012). Offline consolidation of memory varies with time in slow wave sleep and can be accelerated by cuing memory reactivations. *Neurobiology of Learning and Memory*, *98*(2), 103–111. https://doi.org/10.1016/j.nlm.2012.07.002
- Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, 11, 114–126. https://doi.org/10.1038/nrn2762
- *Diekelmann, S., Born, J., & Wagner, U. (2010). Sleep enhances false memories depending on general memory performance. *Behavioural Brain Research*, 208(2), 425–429. https://doi.org/10.1016/j.bbr.2009.12.021

- *Diekelmann, S., Büchel, C., Born, J., & Rasch, B. (2011). Labile or stable: Opposing consequences for memory when reactivated during waking and sleep. *Nature Neuroscience*, *14*(3), 381–386. https://doi.org/10.1038/nn.2744
- *Diekelmann, S., Landolt, H.-P., Lahl, O., Born, J., & Wagner, U. (2008). Sleep loss produces false memories. *PLOS ONE*, *3*(10), Article e3512. https://doi.org/10.1371/journal.pone.0003512
- 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
- *Donohue, K. C., & Spencer, R. M. C. (2011). Continuous re-exposure to environmental sound cues during sleep does not improve memory for semantically unrelated word pairs. *Journal of Cognitive Education and Psychology*, *10*(2), 167–177. https://doi.org/10.1891/1945-8959.10.2.167
- *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
- *Drosopoulos, S., Wagner, U., & Born, J. (2005). Sleep enhances explicit recollection in recognition memory. *Learning & Memory*, *12*, 44–51. https://doi.org/10.1101/lm.83805
- *Drosopoulos, S., Windau, E., Wagner, U., & Born, J. (2007). Sleep enforces the temporal order in memory. *PLOS ONE*, *2*(4), Article e376. https://doi.org/10.1371/journal.pone.0000376
- 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
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*, 629–634. https://doi.org/10.1136/bmj.315.7109.629
- Ego-Stengel, V., & Wilson, M. A. (2010). Disruption of ripple-associated hippocampal activity during rest impairs spatial learning in the rat. *Hippocampus*, *20*(1), 1–10. https://doi.org/10.1002/hipo.20707
- *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., Hulbert, J. C., Stickgold, R., Dinges, D. F., & Thompson-Schill, S. L. (2006).

 Interfering with theories of sleep and memory: Sleep, declarative memory, and associative interference. *Current Biology*, *16*(13), 1290–1294. https://doi.org/10.1016/j.cub.2006.05.024
- 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
- Ellis, P. D. (2010). *The essential guide to effect sizes*. Cambridge University Press. https://doi.org/10.1017/CBO9780511761676

- 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., & Heck, D. W. (2019). Detecting evidential value and p-hacking with the p-curve tool: A word of caution. *Zeitschrift für Psychologie*, 227(4), 249–260. https://doi.org/10.1027/2151-2604/a000383
- 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
- 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
- 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

- *Feld, G. B., Weis, P. P., & Born, J. (2016). The limited capacity of sleep-dependent memory consolidation. *Frontiers in Psychology*, *7*, Article 1368. https://doi.org/10.3389/fpsyg.2016.01368
- *Fenn, K. M. (2006). Waking up to the impact of sleep: Consolidation of generalized skill learning and declarative memory formation (Publication No. 3231393) [Doctoral dissertation, University of Chicago]. ProQuest Dissertations and Theses Global.
- *Fenn, K. M., Gallo, D. A., Margoliash, D., Roediger, H. L., III, & Nusbaum, H. C. (2009). Reduced false memory after sleep. *Learning & Memory*, *16*, 509–513. https://doi.org/10.1101/lm.1500808
- *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
- Fenn, K. M., & Hambrick, D. Z. (2015). General intelligence predicts memory change across sleep.

 *Psychonomic Bulletin & Review, 22, 791–799. https://doi.org/10.3758/s13423-014-0731-1
- 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
- *Fischer, S., Diekelmann, S., & Born, J. (2011). Sleep's role in the processing of unwanted memories.

 *Journal of Sleep Research, 20(2), 267–274. https://doi.org/10.1111/j.1365-2869.2010.00881.x
- *Fogler, K. (2011). The effects of episodic gist and sleep on false memory (Publication No. 3465472)

 [Doctoral dissertation, Saint Louis University]. ProQuest Dissertations and Theses Global.
- Foster, D. J., & Wilson, M. A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, *440*, 680–683. https://doi.org/10.1038/nature04587
- *Fowler, M. J., Sullivan, M. J., & Ekstrand, B. R. (1973). Sleep and memory. *Science*, *179*(4070), 302–304. https://doi.org/10.1126/science.179.4070.302

- 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., Albouy, G., Boly, M., Dang-Vu, T. T., Darsaud, A., Desseilles, M., Rauchs, G., Schabus, M., Sterpenich, V., Vandewalle, G., Maquet, P., & Peigneux, P. (2007). Sleep transforms the cerebral trace of declarative memories. *Proceedings of the National Academy of Sciences of the United States of America*, 104(47), 18778–18783. https://doi.org/10.1073/pnas.0705454104
- *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
- *Gais, S., Rasch, B., Dahmen, J. C., Sara, S., & Born, J. (2011). The memory function of noradrenergic activity in non-REM sleep. *Journal of Cognitive Neuroscience*, *23*(9), 2582–2592. https://doi.org/10.1162/jocn.2011.21622
- *Gais, S., Sommer, M., Fischer, S., Perras, B., & Born, J. (2002). Post-trial administration of vasopressin in humans does not enhance memory formation (vasopressin and memory consolidation). *Peptides, 23*(3), 581–583. https://doi.org/10.1016/S0196-9781(01)00625-8
- Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *irr: Various coefficients of interrater reliability* and agreement (Version 0.84.1) [R package]. https://cran.r-project.org/package=irr
- *Genzel, L., Bäurle, A., Potyka, A., Wehrle, R., Adamczyk, M., Friess, E., Steiger, A., & Dresler, M. (2014). Diminished nap effects on memory consolidation are seen under oral contraceptive use.

 *Neuropsychobiology, 70(4), 253–261. https://doi.org/10.1159/000369022
- *Genzel, L., Kiefer, T., Renner, L., Wehrle, R., Kluge, M., Grözinger, M., Steiger, A., & Dresler, M. (2012). Sex and modulatory menstrual cycle effects on sleep related memory consolidation.

 *Psychoneuroendocrinology, 37(7), 987–998. https://doi.org/10.1016/j.psyneuen.2011.11.006

- *Giganti, F., Arzilli, C., Conte, F., Toselli, M., Viggiano, M. P., & Ficca, G. (2014). The effect of daytime nap on priming and recognition tasks in preschool children. *Sleep*, *37*(6), 1087–1093. https://doi.org/10.5665/sleep.3766
- Girardeau, G., Benchenane, K., Wiener, S. I., Buzsáki, G., & Zugaro, M. B. (2009). Selective suppression of hippocampal ripples impairs spatial memory. *Nature Neuroscience*, *12*(10), 1222–1223. https://doi.org/10.1038/nn.2384
- Girardeau, G., & Zugaro, M. (2011). Hippocampal ripples and memory consolidation. *Current Opinion in Neurobiology*, *21*(3), 452–459. https://doi.org/10.1016/j.conb.2011.02.005
- 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
- *Göder, R., Graf, A., Ballhausen, F., Weinhold, S., Baier, P. C., Junghanns, K., & Prehn-Kristensen, A. (2015). Impairment of sleep-related memory consolidation in schizophrenia: Relevance of sleep spindles? *Sleep Medicine*, *16*(5), 564–569. https://doi.org/10.1016/j.sleep.2014.12.022
- 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
- Goldstein, A. N., & Walker, M. P. (2014). The role of sleep in emotional brain function. *Annual Review of Clinical Psychology*, *10*, 679–708. https://doi.org/10.1146/annurev-clinpsy-032813-153716
- *Gorfine, T., Yeshurun, Y., & Zisapel, N. (2007). Nap and melatonin-induced changes in hippocampal activation and their role in verbal memory consolidation. *Journal of Pineal Research*, *43*(4), 336–342. https://doi.org/10.1111/j.1600-079X.2007.00482.x

- *Griessenberger, H., Hoedlmoser, K., Heib, D. P. J., Lechinger, J., Klimesch, W., & Schabus, M. (2012).

 Consolidation of temporal order in episodic memories. *Biological Psychology*, *91*(1), 150–155.

 https://doi.org/10.1016/j.biopsycho.2012.05.012
- Gu, X., Hoijtink, H., Mulder, J., van Lissa, C. J., van Zundert, C., Jones, J. & Waller, N. (2020). bain:

 Bayes factors for informative hypotheses (Version 0.2.4) [R package]. https://CRAN.R-project.org/package=bain
- 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
- 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
- Hamann, S. B., Ely, T. D., Hoffman, J. M., & Kilts, C. D. (2002). Ecstasy and agony: Activation of the human amygdala in positive and negative emotion. *Psychological Science*, *13*(2), 135–141. https://doi.org/10.1111/1467-9280.00425
- Han, Z., Yang, Y., Zhang, Q., & Mo, L. (2020). Can you voluntarily forget what you are planning to forget? Behavioral evidence for the underlying truth of the cost–benefit principle. *Psychological Research*. Advance online publication. https://doi.org/10.1007/s00426-020-01339-8
- *Hanert, A., Weber, F. D., Pedersen, A., Born, J., & Bartsch, T. (2017). Sleep in humans stabilizes pattern separation performance. *The Journal of Neuroscience*, *37*(50), 12238–12246. https://doi.org/10.1523/JNEUROSCI.1189-17.2017
- *Havas, V., Taylor, J. S. H., Vaquero, L., de Diego-Balaguer, R., Rodríguez-Fornells, A., & Davis, M. H. (2018). Semantic and phonological schema influence spoken word learning and overnight

- consolidation. *Quarterly Journal of Experimental Psychology*, *71*(6), 1469–1481. https://doi.org/10.1080/17470218.2017.1329325
- Hebb, D. O. (1949). The organization of behavior: A neuropsychological theory. Wiley.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators.

 **Journal of Educational Statistics, 6(2), 107–128. https://www.jstor.org/stable/1164588
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. https://doi.org/10.1002/jrsm.5
- *Heim, S., Klann, J., Schattka, K. I., Bauhoff, S., Borcherding, G., Nosbüsch, N., Struth, L., Binkofski, F. C., & Werner, C. J. (2017). A nap but not rest or activity consolidates language learning. *Frontiers in Psychology*, *8*, Article 665. https://doi.org/10.3389/fpsyg.2017.00665
- Helfrich, R. F., Lendner, J. D., Mander, B. A., Guillen, H., Paff, M., Mnatsakanyan, L., Vadera, S., Walker, M. P., Lin, J. J., & Knight, R. T. (2019). Bidirectional prefrontal-hippocampal dynamics organize information transfer during sleep in humans. *Nature Communications, 10*, Article 3572. https://doi.org/10.1038/s41467-019-11444-x
- *Henderson, L. M., Weighall, A. R., Brown, H., & Gaskell, M. G. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science*, *15*(5), 674–687. https://doi.org/10.1111/j.1467-7687.2012.01172.x
- 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
- *Henry, M., Ross, I. L., Wolf, P. S. A., & Thomas, K. G. F. (2017). Impaired quality and efficiency of sleep impairs cognitive functioning in Addison's disease. *Psychoneuroendocrinology*, *78*, 237–245. https://doi.org/10.1016/j.psyneuen.2017.02.004
- *Herzog, N., Friedrich, A., Fujita, N., Gais, S., Jauch-Chara, K., Oltmanns, K. M., & Benedict, C. (2012).

 Effects of daytime food intake on memory consolidation during sleep or sleep deprivation. *PLOS*ONE, 7(6), Article e40298. https://doi.org/10.1371/journal.pone.0040298

- *Himmer, L., Müller, E., Gais, S., & Schönauer, M. (2017). Sleep-mediated memory consolidation depends on the level of integration at encoding. *Neurobiology of Learning and Memory*, *137*, 101–106. https://doi.org/10.1016/j.nlm.2016.11.019
- Himmer, L., Schönauer, M., Heib, D. P. J., Schabus, M., & Gais, S. (2019). Rehearsal initiates systems memory consolidation, sleep makes it last. *Science Advances*, *5*(4), Article eaav1695. https://doi.org/10.1126/sciadv.aav1695
- Hoijtink, H., Mulder, J., van Lissa, C., & Gu, X. (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods*, *24*(5), 549–556. http://dx.doi.org/10.1037/met0000201
- 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., 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
- Humiston, G. B., Tucker, M. A., Summer, T., & Wamsley, E. J. (2019). Resting states and memory consolidation: A preregistered replication and meta-analysis. *Scientific Reports, 9*, Article 19345. https://doi.org/10.1038/s41598-019-56033-6
- *Hupbach, A. (2018). Long-term effects of directed forgetting. *Memory*, *26*(3), 321–329. https://doi.org/10.1080/09658211.2017.1358748
- Hutchison, I. C., & Rathore, S. (2015). The role of REM sleep theta activity in emotional memory.

 Frontiers in Psychology, 6, Article 1439. https://doi.org/10.3389/fpsyg.2015.01439
- *Idzikowski, C. (1984). Sleep and memory. *British Journal of Psychology*, 75(4), 439–449. https://doi.org/10.1111/j.2044-8295.1984.tb01914.x

- *Igloi, K., Gaggioni, G., Sterpenich, V., & Schwartz, S. (2015). A nap to recap or how reward regulates hippocampal-prefrontal memory networks during daytime sleep in humans. *eLife*, *4*, Article e07903. https://doi.org/10.7554/elife.07903
- Iglowstein, I., Jenni, O. G., Molinari, L., & Largo, R. H. (2003). Sleep duration from infancy to adolescence: Reference values and generational trends. *Pediatrics*, *111*(2), 302–307. https://doi.org/10.1542/peds.111.2.302
- 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
- Jadhav, S. P., Kemere, C., German, P. W., & Frank, L. M. (2012). Awake hippocampal sharp-wave ripples support spatial memory. *Science*, *336*(6087), 1454–1458. https://doi.org/10.1126/science.1217230
- *James, J. R. (2014). The effect of sleep-dependent memory consolidation on pattern separation and pattern completion in delayed retrieval (Publication No. 3622068) [Doctoral dissertation, Brigham Young University]. ProQuest Dissertations and Theses Global.
- 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
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953
- Johnson, B. T. (2021). Toward a more transparent, rigorous, and generative psychology [Editorial].

 *Psychological Bulletin, 147(1), 1–15. http://dx.doi.org/10.1037/bul0000317
- Johnson, B. T., & Huedo-Medina, T. B. (2011). Depicting estimates using the intercept in metaregression models: The moving constant technique. *Research Synthesis Methods*, 2(3), 204–220. https://doi.org/10.1002/jrsm.49

- *Jones, B. J., Schultz, K. S., Adams, S., Baran, B., & Spencer, R. M. C. (2016). Emotional bias of sleep-dependent processing shifts from negative to positive with aging. *Neurobiology of Aging*, *45*, 178–189. https://doi.org/10.1016/j.neurobiolaging.2016.05.019
- *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
- Kitamura, T., Ogawa, S. K., Roy, D. S., Okuyama, T., Morrissey, M. D., Smith, L. M., Redondo, R. L., & Tonegawa, S. (2017). Engrams and circuits crucial for systems consolidation of a memory. *Science*, *356*(6333), 73–78. https://doi.org/10.1126/science.aam6808
- 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
- *Klinzing, J. G., Rasch, B., Born, J., & Diekelmann, S. (2016). Sleep's role in the reconsolidation of declarative memories. *Neurobiology of Learning and Memory*, *136*, 166–173. https://doi.org/10.1016/j.nlm.2016.10.004
- *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
- 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
- Krippendorff, K. (1980). *Content analysis: An introduction to its methodology* (1st ed.). SAGE Publications.
- *Kurdziel, L., Duclos, K., & Spencer, R. M. C. (2013). Sleep spindles in midday naps enhance learning in preschool children. *Proceedings of the National Academy of Sciences of the United States of America*, 110(43), 17267–17272. https://doi.org/10.1073/pnas.1306418110

- *Kurdziel, L. B. F., & Spencer, R. M. C. (2016). Consolidation of novel word learning in native English-speaking adults. *Memory*, 24(4), 471–481. https://doi.org/10.1080/09658211.2015.1019889
- *Kuriyama, K., Honma, M., Yoshiike, T., & Kim, Y. (2013). Memory suppression trades prolonged fear and sleep-dependent fear plasticity for the avoidance of current fear. *Scientific Reports, 3*, Article 2227. https://doi.org/10.1038/srep02227
- *Lahl, O., Wispel, C., Willigens, B., & Pietrowsky, R. (2008). An ultra short episode of sleep is sufficient to promote declarative memory performance. *Journal of Sleep Research*, *17*(1), 3–10. https://doi.org/10.1111/j.1365-2869.2008.00622.x
- Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology*, 44(7), 701–710. https://doi.org/10.1002/ejsp.2023
- Lakens, D., & Evers, E. R. K. (2014). Sailing from the seas of chaos into the corridor of stability:

 Practical recommendations to increase the informational value of studies. *Perspectives on Psychological Science*, *9*(3), 278–292. https://doi.org/10.1177/1745691614528520
- Lane, R. D., Reiman, E. M., Bradley, M. M., Lang, P. J., Ahern, G. L., Davidson, R. J., & Schwartz, G. E. (1997). Neuroanatomical correlates of pleasant and unpleasant emotion. *Neuropsychologia*, *35*(11), 1437–1444. https://doi.org/10.1016/S0028-3932(97)00070-5
- *Lau, H. (2011). *Daytime napping: Effects on relational memory* (Publication No. 3469885) [Doctoral dissertation, City University of New York]. ProQuest Dissertations and Theses Global.
- *Lau, H., Tucker, M. A., & Fishbein, W. (2010). Daytime napping: Effects on human direct associative and relational memory. *Neurobiology of Learning and Memory*, *93*(4), 554–560. https://doi.org/10.1016/j.nlm.2010.02.003
- Lee, A. K., & Wilson, M. A. (2002). Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron*, *36*(6), 1183–1194. https://doi.org/10.1016/S0896-6273(02)01096-6
- *Lehmann, M., Schreiner, T., Seifritz, E., & Rasch, B. (2016). Emotional arousal modulates oscillatory correlates of targeted memory reactivation during NREM, but not REM sleep. *Scientific Reports*, 6, Article 39229. https://doi.org/10.1038/srep39229

- *Lemos, N., Weissheimer, J., & Ribeiro, S. (2014). Naps in school can enhance the duration of declarative memories learned by adolescents. *Frontiers in Systems Neuroscience*, 8, Article 103. https://doi.org/10.3389/fnsys.2014.00103
- 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
- *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
- *Lin, C.-C., & Yang, C.-M. (2014). Evidence of sleep-facilitating effect on formation of novel semantic associations: An event-related potential (ERP) study. *Neurobiology of Learning and Memory*, *116*, 69–78. https://doi.org/10.1016/j.nlm.2014.08.011
- Lipinska, G., Stuart, B., Thomas, K. G. F., Baldwin, D. S., & Bolinger, E. (2019). Preferential consolidation of emotional memory during sleep: A meta-analysis. *Frontiers in Psychology*, *10*, Article 1014. https://doi.org/10.3389/fpsyg.2019.01014
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE Publications.
- *Lo, J. C., Dijk, D.-J., & Groeger, J. A. (2014). Comparing the effects of nocturnal sleep and daytime napping on declarative memory consolidation. *PLOS ONE*, *9*(9), Article e108100. https://doi.org/10.1371/journal.pone.0108100
- 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

- *Lo, J. C., Sim, S. K. Y., & Chee, M. W. L. (2014). Sleep reduces false memory in healthy older adults. *Sleep*, *37*(4), 665–671. https://doi.org/10.5665/sleep.3564
- Lowe, C. J., Safati, A., & Hall, P. A. (2017). The neurocognitive consequences of sleep restriction: A meta-analytic review. *Neuroscience & Biobehavioral Reviews, 80,* 586–604. https://doi.org/10.1016/j.neubiorev.2017.07.010
- *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
- MacDonald, K. J., & Cote, K. A. (2021). Contributions of post-learning REM and NREM sleep to memory retrieval. *Sleep Medicine Reviews*, *59*, Article 101453. https://doi.org/10.1016/j.smrv.2021.101453
- 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
- *Mantua, J., Mahan, K. M., Henry, O. S., Spencer, R. M. C. (2015). Altered sleep composition after traumatic brain injury does not affect declarative sleep-dependent memory consolidation.

 Frontiers in Human Neuroscience, 9, Article 328. https://doi.org/10.3389/fnhum.2015.00328
- *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
- *Maski, K., Holbrook, H., Manoach, D., Hanson, E., Kapur, K., & Stickgold, R. (2015). Sleep dependent memory consolidation in children with autism spectrum disorder. *Sleep*, *38*(12), 1955–1963. https://doi.org/10.5665/sleep.5248
- *Maski, K., Steinhart, E., Holbrook, H., Katz, E. S., Kapur, K., & Stickgold, R. (2017). Impaired memory consolidation in children with obstructive sleep disordered breathing. *PLOS ONE*, *12*(11), Article e0186915. https://doi.org/10.1371/journal.pone.0186915

- Maurer, L., Zitting, K.-M., Elliott, K., Czeisler, C. A., Ronda, J. M., & Duffy, J. F. (2015). A new face of sleep: The impact of post-learning sleep on recognition memory for face-name associations.

 Neurobiology of Learning and Memory, 126, 31–38. https://doi.org/10.1016/j.nlm.2015.10.012
- *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
- *Mazza, S., Gerbier, E., Gustin, M.-P., Kasikci, Z., Koenig, O., Toppino, T. C., & Magnin, M. (2016).

 Relearn faster and retain longer: Along with practice, sleep makes perfect. *Psychological Science*, 27(10), 1321–1330. https://doi.org/10.1177/0956797616659930
- *McDevitt, E. A., Rowe, K. M., Brady, M., Duggan, K. A., & Mednick, S. C. (2014). The benefit of offline sleep and wake for novel object recognition. *Experimental Brain Research*, *232*, 1487–1496. https://doi.org/10.1007/s00221-014-3830-3
- 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
- McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review*, *39*(4), 352–370. https://doi.org/10.1037/h0069819
- *McKeon, S., Pace-Schott, E. F., & Spencer, R. M. C. (2012). Interaction of sleep and emotional content on the production of false memories. *PLOS ONE*, *7*(11), Article e49353. https://doi.org/10.1371/journal.pone.0049353
- McKinney, F. (1935). Retroactive inhibition and recognition memory. *Journal of Experimental Psychology*, *18*(5), 585–598. https://doi.org/10.1037/h0054474
- McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in metaanalysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science*, *11*(5), 730–749. https://doi.org/10.1177/1745691616662243

- *Mednick, S. C., Cai, D. J., Kanady, J., & Drummond, S. P. A. (2008). Comparing the benefits of caffeine, naps and placebo on verbal, motor and perceptual memory. *Behavioural Brain Research*, 193(1), 79–86. https://doi.org/10.1016/j.bbr.2008.04.028
- 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
- *Meléndez, J., Galli, I., Boric, K., Ortega, A., Zuñiga, L., Henríquez-Roldán, C. F., & Cárdenas, A. M. (2005). Zolpidem and triazolam do not affect the nocturnal sleep-induced memory improvement. *Psychopharmacology, 181, 21–26. https://doi.org/10.1007/s00213-005-2228-0
- *Mograss, M., Godbout, R., & Guillem, F. (2006). The ERP old-new effect: A useful indicator in studying the effects of sleep on memory retrieval processes. *Sleep*, *29*(11), 1491–1500. https://doi.org/10.1093/sleep/29.11.1491
- *Mograss, M. A., Guillem, F., Brazzini-Poisson, V., & Godbout, R. (2009). The effects of total sleep deprivation on recognition memory processes: A study of event-related potential. *Neurobiology of Learning and Memory*, *91*(4), 343–352. https://doi.org/10.1016/j.nlm.2009.01.008
- *Mograss, M. A., Guillem, F., & Godbout, R. (2008). Event-related potentials differentiates the processes involved in the effects of sleep on recognition memory. *Psychophysiology*, *45*(3), 420–434. https://doi.org/10.1111/j.1469-8986.2007.00643.x
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Medicine*, *6*(7), Article e1000097. https://doi.org/10.1371/journal.pmed.1000097
- *Monaghan, P., Shaw, J. J., Ashworth-Lord, A., & Newbury, C. R. (2017). Hemispheric processing of memory is affected by sleep. *Brain and Language*, *167*, 36–43. https://doi.org/10.1016/j.bandl.2016.05.003

- *Morgenthaler, J., Wiesner, C. D., Hinze, K., Abels, L. C., Prehn-Kristensen, A., & Göder, R. (2014).

 Selective REM-sleep deprivation does not diminish emotional memory consolidation in young healthy subjects. *PLOS ONE*, *9*(2), Article e89849. https://doi.org/10.1371/journal.pone.0089849
- 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
- Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs.

 Organizational Research Methods, 11(2), 364–386. https://doi.org/10.1177/1094428106291059
- Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*(1), 105–125. https://doi.org/10.1037//1082-989X.7.1.105
- 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
- Müller, G. E., & Pilzecker, A. (1900). Experimentelle Beiträge zur Lehre vom Gedächtnis

 [Experimental contributions to the theory of memory]. Zeitschrift für Psychologie und Physiologie

 der Sinnesorgane, Ergänzungsband [Supplement] 1, 1–300.
- Mundt, D., Abel, R., & Hänze, M. (2020). Exploring the effect of testing on forgetting in vocabulary learning: An examination of the bifurcation model. *Journal of Cognitive Psychology*, *32*(2), 214–228. https://doi.org/10.1080/20445911.2020.1733584
- 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
- *Nesca, M., & Koulack, D. (1994). Recognition memory, sleep and circadian rhythms. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 48(3), 359–379. https://doi.org/10.1037/1196-1961.48.3.359

- 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
- *Nishida, M., Pearsall, J., Buckner, R. L., & Walker, M. P. (2009). REM sleep, prefrontal theta, and the consolidation of human emotional memory. *Cerebral Cortex*, *19*(5), 1158–1166. https://doi.org/10.1093/cercor/bhn155
- *Nissen, C., Kloepfer, C., Feige, B., Piosczyk, H., Spiegelhalder, K., Voderholzer, U., & Riemann, D. (2011). Sleep-related memory consolidation in primary insomnia. *Journal of Sleep Research*, 20(1), 129–136. https://doi.org/10.1111/j.1365-2869.2010.00872.x
- Nissen, C., Piosczyk, H., Holz, J., Maier, J. G., Frase, L., Sterr, A., Riemann, D., & Feige, B. (2021). Sleep is more than rest for plasticity in the human cortex. *Sleep, 44*(3), Article zsaa216. https://doi.org/10.1093/sleep/zsaa216
- Noldy, N. E., Stelmack, R. M., & Campbell, K. B. (1990). Event-related potentials and recognition memory for pictures and words: The effects of intentional and incidental learning.

 Psychophysiology, 27(4), 417–428. https://doi.org/10.1111/j.1469-8986.1990.tb02337.x
- Ohayon, M. M., Carskadon, M. A., Guilleminault, C., & Vitiello, M. V. (2004). Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: Developing normative sleep values across the human lifespan. *Sleep*, *27*(7), 1255–1273. https://doi.org/10.1093/sleep/27.7.1255
- 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
- *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

- *Oyarzún, J. P., Morís, J., Luque, D., de Diego-Balaguer, R., & Fuentemilla, L. (2017). Targeted memory reactivation during sleep adaptively promotes the strengthening or weakening of overlapping memories. *Journal of Neuroscience*, *37*(32), 7748–7758. https://doi.org/10.1523/JNEUROSCI.3537-16.2017
- 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
- Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 3–8. https://doi.org/10.1037/0278-7393.31.1.3
- Pastötter, B., & Bäuml, K.-H. T. (2016). Reversing the testing effect by feedback: Behavioral and electrophysiological evidence. *Cognitive, Affective, & Behavioral Neuroscience*, *16*, 473–488. https://doi.org/10.3758/s13415-016-0407-6
- *Pawlizki, A. (2012). Spezifikation des deklarativen Schlafeffekts und Signaturen von deklarativem Gedächtnis im Elektroenzephalogramm [Specification of the declarative sleep effect and signatures of declarative memory in the electroencephalogram; Doctoral dissertation, Ludwig Maximilian University Munich]. Electronic Theses of the Ludwig Maximilian University Library. https://nbn-resolving.de/urn:nbn:de:bvb:19-153665

- 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. (2011). Sleep leads to changes in the emotional memory trace:

 Evidence from fMRI. *Journal of Cognitive Neuroscience*, *23*(6), 1285–1297.

 https://doi.org/10.1162/jocn.2010.21526
- *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., Schacter, D. L., Propper, R. E., Huang, L.-W., Wamsley, E. J., Tucker, M. A., Walker, M. P., & Stickgold, R. (2009). The role of sleep in false memory formation. *Neurobiology of Learning* and *Memory*, 92(3), 327–334. https://doi.org/10.1016/j.nlm.2009.03.007
- *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
- 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
- Pfeiffer, B. E. (2020). The content of hippocampal "replay". *Hippocampus*, *30*(1), 6–18. https://doi.org/10.1002/hipo.22824
- *Piosczyk, H., Holz, J., Feige, B., Spiegelhalder, K., Weber, F., Landmann, N., Kuhn, M., Frase, L., Riemann, D., Voderholzer, U., & Nissen, C. (2013). The effect of sleep-specific brain activity versus reduced stimulus interference on declarative memory consolidation. *Journal of Sleep Research*, 22(4), 406–413. https://doi.org/10.1111/jsr.12033

- *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
- Popper, K. (2002). *The logic of scientific discovery* (2nd ed.). Routledge. https://doi.org/10.4324/9780203994627
- Postman, L. (1952). Retroactive inhibition in recall and recognition. *Journal of Experimental Psychology*, 44(3), 165–169. https://doi.org/10.1037/h0057388
- *Potkin, K. T., & Bunney, W. E., Jr. (2012). Sleep improves memory: The effect of sleep on long term memory in early adolescence. *PLOS ONE*, *7*(8), Article e42191. https://doi.org/10.1371/journal.pone.0042191
- *Prehn-Kristensen, A., Göder, R., Chirobeja, S., Breßmann, I., Ferstl, R., & Baving, L. (2009). Sleep in children enhances preferentially emotional declarative but not procedural memories. *Journal of Experimental Child Psychology*, 104(1), 132–139. https://doi.org/10.1016/j.jecp.2009.01.005
- *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., Lotzkat, K., Bauhofer, E., Wiesner, C. D., & Baving, L. (2015). Sleep supports memory of odors in adults but not in children. *PLOS ONE, 10*(9), Article e0139069. https://doi.org/10.1371/journal.pone.0139069
- *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
- *Prehn-Kristensen, A., Munz, M., Molzow, I., Wilhelm, I., Wiesner, C. D., & Baving, L. (2013). Sleep promotes consolidation of emotional memory in healthy children but not in children with attention-deficit hyperactivity disorder. *PLOS ONE*, *8*(5), Article e65098. https://doi.org/10.1371/journal.pone.0065098
- Puentes-Mestril, C., & Aton, S. J. (2017). Linking network activity to synaptic plasticity during sleep:

 Hypotheses and recent data. *Frontiers in Neural Circuits*, *11*, Article 61.

 https://doi.org/10.3389/fncir.2017.00061
- Pustejovsky, J. E. (2020). *clubSandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections* (Version 0.4.2) [R package]. https://cran.r-project.org/package=clubSandwich
- 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
- Pustejovsky, J. E., & Tipton, E. (2021). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science*. Advance online publication. https://doi.org/10.1007/s11121-021-01246-3
- *Racsmány, M., Conway, M. A., & Demeter, G. (2010). Consolidation of episodic memories during sleep: Long-term effects of retrieval practice. *Psychological Science*, *21*(1), 80–85. https://doi.org/10.1177/0956797609354074
- 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
- 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
- *Renfro, A. G. (2015). What happened last night? Sleep, sex, and recollection (Publication No. 1595100) [Master's thesis, Eastern Kentucky University]. ProQuest Dissertations and Theses Global.
- 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
- 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
- Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003
- Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803
- Rohatgi, A. (2019). WebPlotDigitizer—Web based plot digitizer (Version 4.1, 4.2) [Web-based computer software]. https://automeris.io/WebPlotDigitizer/
- 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://dx.doi.org/10.1037/a0037559
- Rummel, J., Marevic, I., & Kuhlmann, B. G. (2016). Investigating storage and retrieval processes of directed forgetting: A model-based approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(10), 1526–1543. https://doi.org/10.1037/xlm0000266

- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178. https://doi.org/10.1037/h0077714
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145
- Russell, J. A. (2009). Emotion, core affect, and psychological construction. *Cognition and Emotion*, 23(7), 1259–1283. https://doi.org/10.1080/02699930902809375
- *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
- 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
- *Schichl, M., Ziberi, M., Lahl, O., & Pietrowsky, R. (2011). The influence of midday naps and relaxation-hypnosis on declarative and procedural memory performance. *Sleep and Hypnosis*, 13(1–2), 7–14. https://www.sleepandhypnosis.org/ing/abstract.aspx?MkID=190
- 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
- Schnuerch, M., & Erdfelder, E. (2020). Controlling decision errors with minimal costs: The sequential probability ratio t test. *Psychological Methods*, *25*(2), 206–226. https://doi.org/10.1037/met0000234
- *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

- *Schönauer, M., Grätsch, M., & Gais, S. (2015). Evidence for two distinct sleep-related long-term memory consolidation processes. *Cortex*, *63*, 68–78. https://doi.org/10.1016/j.cortex.2014.08.005
- *Schönauer, M., Pawlizki, A., Köck, C., & Gais, S. (2014). Exploring the effect of sleep and reduced interference on different forms of declarative memory. *Sleep*, *37*(12), 1995–2007. https://doi.org/10.5665/sleep.4258
- Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*(2), 322–339. https://doi.org/10.1037/met0000061
- 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
- *Schreiner, T., & Rasch, B. (2015). Boosting vocabulary learning by verbal cueing during sleep.

 *Cerebral Cortex, 25(11), 4169–4179. https://doi.org/10.1093/cercor/bhu139
- 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
- *Scullin, M. K., Fairley, J., Decker, M. J., & Bliwise, D. L. (2017). The effects of an afternoon nap on episodic memory in young and older adults. *Sleep*, *40*(5), Article zsx035. https://doi.org/10.1093/sleep/zsx035
- *Scullin, M. K., & McDaniel, M. A. (2010). Remembering to execute a goal: Sleep on it! *Psychological Science*, *21*(7), 1028–1035. https://doi.org/10.1177/0956797610373373
- Sekeres, M. J., Winocur, G., Moscovitch, M., Anderson, J. A. E., Pishdadian, S., Wojtowicz, J. M., St-Laurent, M., McAndrews, M. P., & Grady, C. L. (2018). Changes in patterns of neural activity

- underlie a time-dependent transformation of memory in rats and humans. *Hippocampus, 28*(10), 745–764. https://doi.org/10.1002/hipo.23009
- Sergerie, K., Chochol, C., & Armony, J. L. (2008). The role of the amygdala in emotional processing: A quantitative meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, 32(4), 811–830. https://doi.org/10.1016/j.neubiorev.2007.12.002
- *Sheth, B. R., Nguyen, N., & Janvelyan, D. (2009). Does sleep really influence face recognition memory? *PLOS ONE*, *4*(5), Article e5496. https://doi.org/10.1371/journal.pone.0005496
- *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
- Siegel, J. M. (2001). The REM sleep-memory consolidation hypothesis. *Science*, *294*(5544), 1058–1063. https://doi.org/10.1126/science.1063049
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. https://doi.org/10.1177/0956797611417632
- Skeldon, A. C., Derks, G., & Dijk, D.-J. (2016). Modelling changes in sleep timing and duration across the lifespan: Changes in circadian rhythmicity or sleep homeostasis? *Sleep Medicine Reviews*, *28*, 96–107. https://doi.org/10.1016/j.smrv.2015.05.011
- Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and metaanalysis. *Psychonomic Bulletin & Review, 8*(2), 203–220. https://doi.org/10.3758/BF03196157
- Sneve, M. H., Grydeland, H., Nyberg, L., Bowles, B., Amlien, I. K., Langnes, E., Walhovd, K. B., & Fjell,
 A. M. (2015). Mechanisms underlying encoding of short-lived versus durable episodic memories.
 The Journal of Neuroscience, 35(13), 5202–5212. https://doi.org/10.1523/JNEUROSCI.4434-14.2015
- *Sonni, A., & Spencer, R. M. C. (2015). Sleep protects memories from interference in older adults.

 Neurobiology of Aging, 36(7), 2272–2281. https://doi.org/10.1016/j.neurobiologing.2015.03.010

- Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. *Psychological Bulletin*, *144*(12), 1325–1346. https://doi.org/10.1037/bul0000169
- Staresina, B. P., Alink, A., Kriegeskorte, N., & Henson, R. N. (2013). Awake reactivation predicts memory in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 110(52), 21159–21164. https://doi.org/10.1073/pnas.1311989110
- Staresina, B. P., Bergmann, T. O., Bonnefond, M., van der Meij, R., Jensen, O., Deuker, L., Elger, C. E., Axmacher, N., & Fell, J. (2015). Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. *Nature Neuroscience*, *18*, 1679–1686. https://doi.org/10.1038/nn.4119
- Staugaard, S. R., & Berntsen, D. (2019). Retrieval intentionality and forgetting: How retention time and cue distinctiveness affect involuntary and voluntary retrieval of episodic memories. *Memory & Cognition*, *47*(5), 893–905. https://doi.org/10.3758/s13421-019-00904-w
- Steriade, M. (2006). Grouping of brain rhythms in corticothalamic systems. *Neuroscience*, *137*(4), 1087–1106. https://doi.org/10.1016/j.neuroscience.2005.10.029
- Sterne, J. A. C., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology*, *54*(10), 1046–1055.

 https://doi.org/10.1016/S0895-4356(01)00377-8
- *Sterpenich, V., Albouy, G., Boly, M., Vandewalle, G., Darsaud, A., Balteau, E., Dang-Vu, T. T.,

 Desseilles, M., D'Argembeau, A., Gais, S., Rauchs, G., Schabus, M., Degueldre, C., Luxen, A.,

 Collette, F., & Maquet, P. (2007). Sleep-related hippocampo-cortical interplay during emotional memory recollection. *PLOS Biology*, *5*(11), Article e282.

 https://doi.org/10.1371/journal.pbio.0050282
- *Sterpenich, V., Ceravolo, L., & Schwartz, S. (2017). Sleep deprivation disrupts the contribution of the hippocampus to the formation of novel lexical associations. *Brain and Language*, *167*, 61–71. https://doi.org/10.1016/j.bandl.2016.12.007

- 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
- 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
- *Studte, S., Bridger, E., & Mecklinger, A. (2015). Nap sleep preserves associative but not item memory performance. *Neurobiology of Learning and Memory*, *120*, 84–93. https://doi.org/10.1016/j.nlm.2015.02.012
- Takashima, A., Nieuwenhuis, I. L. C., Jensen, O., Talamini, L. M., Rijpkema, M., & Fernández, G.
 (2009). Shift from hippocampal to neocortical centered retrieval network with consolidation. *The Journal of Neuroscience*, 29(32), 10087–10093. https://doi.org/10.1523/JNEUROSCI.0799-09.2009
- 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. *Proceedings of the National Academy of Sciences of the United States of America*, 103(3), 756–761. https://doi.org/10.1073/pnas.0507774103
- *Takeuchi, M., Furuta, H., Sumiyoshi, T., Suzuki, M., Ochiai, Y., Hosokawa, M., Matsui, M., & Kurachi, M. (2014). Does sleep improve memory organization? *Frontiers in Behavioral Neuroscience, 8*, Article 65. https://doi.org/10.3389/fnbeh.2014.00065
- *Talamini, L. M., Nieuwenhuis, I. L. C., Takashima, A., & Jensen, O. (2008). Sleep directly following learning benefits consolidation of spatial associative memory. *Learning & Memory*, *15*, 233–237. https://doi.org/10.1101/lm.771608

- Tambini, A., Ketz, N., & Davachi, L. (2010). Enhanced brain correlations during rest are related to memory for recent experiences. *Neuron*, 65(2), 280–290. https://doi.org/10.1016/j.neuron.2010.01.001
- *Tamminen, J., Lambon Ralph, M. A., & Lewis, P. A. (2017). Targeted memory reactivation of newly learned words during sleep triggers REM-mediated integration of new memories and existing knowledge. *Neurobiology of Learning and Memory*, *137*, 77–82. https://doi.org/10.1016/j.nlm.2016.11.012
- *Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *Journal of Neuroscience*, *30*(43), 14356–14360. https://doi.org/10.1523/JNEUROSCI.3028-10.2010
- Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology*, *2*, 85–112. https://doi.org/10.1007/s40865-016-0026-5
- *Tempesta, D., De Gennaro, L., Natale, V., & Ferrara, M. (2015). Emotional memory processing is influenced by sleep quality. *Sleep Medicine*, *16*(7), 862–870. https://doi.org/10.1016/j.sleep.2015.01.024
- Tempesta, D., Socci, V., De Gennaro, L., & Ferrara, M. (2018). Sleep and emotional processing. *Sleep Medicine Reviews*, 40, 183–195. https://doi.org/10.1016/j.smrv.2017.12.005
- *Tempesta, D., Socci, V., Dello Ioio, G., De Gennaro, L., & Ferrara, M. (2017). The effect of sleep deprivation on retrieval of emotional memory: A behavioural study using film stimuli.

 Experimental Brain Research, 235, 3059–3067. https://doi.org/10.1007/s00221-017-5043-z
- Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression.

 *Psychological Methods, 20(3), 375–393. https://doi.org/10.1037/met0000011
- Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, 40(6), 604–634. https://doi.org/10.3102/1076998615606099

- Tompary, A., Duncan, K., & Davachi, L. (2015). Consolidation of associative and item memory is related to post-encoding functional connectivity between the ventral tegmental area and different medial temporal lobe subregions during an unrelated task. *The Journal of Neuroscience*, 35(19), 7326–7331. https://doi.org/10.1523/JNEUROSCI.4816-14.2015
- 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
- Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., Witter, M. P., & Morris, R. G. M. (2007). Schemas and memory consolidation. *Science*, *316*(5821), 76–82. https://doi.org/10.1126/science.1135935
- *Tucker, M. A., Hirota, Y., Wamsley, E. J., Lau, H., Chaklader, A., & Fishbein, W. (2006). A daytime nap containing solely non-REM sleep enhances declarative but not procedural memory.

 *Neurobiology of Learning and Memory, 86(2), 241–247.

 https://doi.org/10.1016/j.nlm.2006.03.005
- *Tucker, M. A., Tang, S. X., Uzoh, A., Morgan, A., & Stickgold, R. (2011). To sleep, to strive, or both:

 How best to optimize memory. *PLOS ONE*, *6*(7), Article e21737.

 https://doi.org/10.1371/journal.pone.0021737
- 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
- *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
- Ulrich, R., Miller, J., & Erdfelder, E. (2018). Effect size estimation from t-statistics in the presence of publication bias: A brief review of existing approaches with some extensions. *Zeitschrift für Psychologie*, 226(1), 56–80. https://doi.org/10.1027/2151-2604/a000319
- Underwood, B. J. (1957). Interference and forgetting. *Psychological Review*, *64*(1), 49–60. https://dx.doi.org/10.1037/h0044616
- *Urbain, C., De Tiège, X., Op De Beeck, M., Bourguignon, M., Wens, V., Verheulpen, D., Van Bogaert, P., & Peigneux, P. (2016). Sleep in children triggers rapid reorganization of memory-related brain processes. *Neurolmage*, *134*, 213–222. https://doi.org/10.1016/j.neuroimage.2016.03.055
- *Urbain, C., Di Vincenzo, T., Peigneux, P., & Van Bogaert, P. (2011). Is sleep-related consolidation impaired in focal idiopathic epilepsies of childhood? A pilot study. *Epilepsy & Behavior*, *22*(2), 380–384. https://doi.org/10.1016/j.yebeh.2011.07.023
- *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
- van der Steen, J. T., van den Bogert, C. A., van Soest-Poortvliet, M. C., Farsani, S. F., Otten, R. H. J., ter Riet, G., & Bouter, L. M. (2018). Determinants of selective reporting: A taxonomy based on content analysis of a random selection of the literature. *PLOS ONE*, *13*(2), Article e0188247. https://doi.org/10.1371/journal.pone.0188247
- van Lissa, C. J., Gu, X., Mulder, J., Rosseel, Y., van Zundert, C., & Hoijtink, H. (2021). Teacher's corner: Evaluating informative hypotheses using the Bayes factor in structural equation models.

- Structural Equation Modeling: A Multidisciplinary Journal, 28(2), 292–301. https://doi.org/10.1080/10705511.2020.1745644
- *van Rijn, E., Carter, N., McMurtrie, H., Willner, P., & Blagrove, M. T. (2017). Sleep does not cause false memories on a story-based test of suggestibility. *Consciousness and Cognition*, *52*, 39–46. https://doi.org/10.1016/j.concog.2017.04.010
- *van Rijn, E., Lucignoli, C., Izura, C., & Blagrove, M. T. (2017). Sleep-dependent memory consolidation is related to perceived value of learned material. *Journal of Sleep Research*, *26*(3), 302–308. https://doi.org/10.1111/jsr.12457
- *van Schalkwijk, F. J., Sauter, C., Hoedlmoser, K., Heib, D. P. J., Klösch, G., Moser, D., Gruber, G., Anderer, P., Zeitlhofer, J., & Schabus, M. (2019). The effect of daytime napping and full-night sleep on the consolidation of declarative and procedural information. *Journal of Sleep Research*, 28(1), Article e12649. https://doi.org/10.1111/jsr.12649
- *Vermeulen, M. C. M., van der Heijden, K. B., Benjamins, J. S., Swaab, H., & van Someren, E. J. W. (2017). Memory effects of sleep, emotional valence, arousal and novelty in children. *Journal of Sleep Research*, 26(3), 309–317. https://doi.org/10.1111/jsr.12506
- 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
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03
- Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis.

 *Research Synthesis Methods, 1(2), 112–125. https://doi.org/10.1002/jrsm.11
- *Wagner, U., Gais, S., & Born, J. (2001). Emotional memory formation is enhanced across sleep intervals with high amounts of rapid eye movement sleep. *Learning & Memory*, 8, 112–119. https://doi.org/10.1101/lm.36801

- *Wagner, U., Kashyap, N., Diekelmann, S., & Born, J. (2007). The impact of post-learning sleep vs. wakefulness on recognition memory for faces with different facial expressions. *Neurobiology of Learning and Memory*, *87*(4), 679–687. https://doi.org/10.1016/j.nlm.2007.01.004
- Walker, M. P. (2009). The role of sleep in cognition and emotion. *Annals of the New York Academy of Sciences*, 1156(1), 168–197. https://doi.org/10.1111/j.1749-6632.2009.04416.x
- 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
- Wamsley, E. J. (2019). Memory consolidation during waking rest. *Trends in Cognitive Sciences*, *23*(3), 171–173. https://doi.org/10.1016/j.tics.2018.12.007
- *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, D.-Y., Liu, D.-Q., Li, S.-F., & Zang, Y.-F. (2012). Increased local synchronization of resting-state fMRI signal after episodic memory encoding reflects off-line memory consolidation. *NeuroReport*, 23(15), 873–878. https://doi.org/10.1097/WNR.0b013e3283587c96
- *Wang, J.-Y., Weber, F. D., Zinke, K., Inostroza, M., & Born, J. (2018). More effective consolidation of episodic long-term memory in children than adults—Unrelated to sleep. *Child Development*, 89(5), 1720–1734. https://doi.org/10.1111/cdev.12839
- *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
- Wang, S. Y., Baker, K. C., Culbreth, J. L., Tracy, O., Arora, M., Liu, T., Morris, S., Collins, M. B., & Wamsley, E. J. (2021). 'Sleep-dependent' memory consolidation? Brief periods of post-training rest and sleep provide an equivalent benefit for both declarative and procedural memory.

 *Learning & Memory, 28, 195–203. https://doi.org/10.1101/lm.053330.120

- Wardle-Pinkston, S., Slavish, D. C., & Taylor, D. J. (2019). Insomnia and cognitive performance: A systematic review and meta-analysis. *Sleep Medicine Reviews*, *48*, Article 101205. https://doi.org/10.1016/j.smrv.2019.07.008
- Weber, F. D., Wang, J.-Y., Born, J., & Inostroza, M. (2014). Sleep benefits in parallel implicit and explicit measures of episodic memory. *Learning & Memory, 21*, 190–198. https://doi.org/10.1101/lm.033530.113
- *Wiesner, C. D., Pulst, J., Krause, F., Elsner, M., Baving, L., Pedersen, A., Prehn-Kristensen, A., & Göder, R. (2015). The effect of selective REM-sleep deprivation on the consolidation and affective evaluation of emotional memories. *Neurobiology of Learning and Memory, 122*, 131–141. https://doi.org/10.1016/j.nlm.2015.02.008
- Wikenheiser, A. M., & Redish, A. D. (2013). The balance of forward and backward hippocampal sequences shifts across behavioral states. *Hippocampus*, *23*(1), 22–29. https://doi.org/10.1002/hipo.22049
- *Wilhelm, I., Diekelmann, S., & Born, J. (2008). Sleep in children improves memory performance on declarative but not procedural tasks. *Learning & Memory*, *15*, 373–377. https://doi.org/10.1101/lm.803708
- *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
- *Wilhelm, I., Wagner, U., & Born, J. (2011). Opposite effects of cortisol on consolidation of temporal sequence memory during waking and sleep. *Journal of Cognitive Neuroscience*, *23*(12), 3703–3712. https://doi.org/10.1162/jocn_a_00093

- *Williams, S. E., & Horst, J. S. (2014). Goodnight book: Sleep consolidation improves word learning via storybooks. *Frontiers in Psychology*, *5*, Article 184. https://doi.org/10.3389/fpsyg.2014.00184
- 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
- *Xi, Z. (2014). Sleep and directed forgetting: A dual-route multinomial tree model (Publication No. 3669640) [Doctoral dissertation, Purdue University]. ProQuest Dissertations and Theses Global.
- 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
- Zhang, H., Fell, J., & Axmacher, N. (2018). Electrophysiological mechanisms of human memory consolidation. *Nature Communications*, *9*, Article 4103. https://doi.org/10.1038/s41467-018-06553-y

Table 1Descriptive Summary of Sample, Study, and Moderator Characteristics

Variable	k	Missing values	М	SD	Mdn	Mode	Range	Krippen- dorff's α
			Sample	characteris	tics			
Sample size	823	0	36.37	29.48	30.00	20.00	7–354	.97
Female ratio	585	238	0.54	0.25	0.56	0.00	0–1	.84
Education	432	391						1
Pre-school children	4							
Pupils	53							
No secondary school certificate	0							
Intermediate school certificate	6							
Qualification for college/university entrance or at least 12 years of education	369							
Chronotype sleep condition	75	748						.39
Morning	0							
Intermediate	75							
Evening	0							
Chronotype wake condition	75	748						.39
Morning	1							
Intermediate	74							
Evening	0							
Country	823	0						1

Variable	k	Missing values	М	SD	Mdn	Mode	Range	Krippen- dorff's α
Austria	4	,						
Australia	4							
Belgium	26							
Brazil	8							
Canada	11							
Switzerland	40							
Chile	2							
China	6							
Germany	297							
Spain	23							
France	31							
UK	64							
Hungary	6							
Israel	1							
Italy	12							
Japan	5							
Netherlands	5							
Poland	2							
Russia	2							
Singapore	3							
Taiwan	2							
USA	268							
South Africa	1							
Year of publication	823	0	2010.38	9.32	2013	2014	1967–2019	1

Variable	k	Missing values	М	SD	Mdn	Mode	Range	Krippen- dorff's α
Whereabouts during the sleep retention interval	803	20		-				.87
Home	339							
Laboratory	448							
Dorm in nursery or school	16							
Whereabouts during the wake retention interval	799	24						.79
Everyday places	464							
Laboratory	335							
Activity during the wake retention interval	790	33						.80
Usual everyday activity	457							
Controlled cognitive activity during whole retention interval	219							
Controlled cognitive activity during part of the retention interval	96							
Wakeful rest	18							
Sleep parameter assessment	753	70						.75
Self-report	225							
Sleep tracker	54							
PSG at home	43							
PSG in laboratory	431							

Variable	k	Missing values	М	SD	Mdn	Mode	Range	Krippen- dorff's α
Sleep duration during wake retention interval	796	27	0.00	0.00	0.00	0.00	0-0.0033	.73
Sleep quality in experimental night	14	809						.37
Below average to average	0							
Above average	14							
Sleepiness learning session	268	555						.92
No significant difference between sleep and wake conditions	235							
Wake condition significantly more sleepy	0							
Sleep condition significantly more sleepy	33							
Sleepiness testing session	272	551						.69
No significant difference between sleep and wake conditions	213							
Wake condition significantly more sleepy	43							
Sleep condition significantly more sleepy	16							
Prestudy sleep, sleep condition	341	482	7.02	1.06	7.10	6.00	4.00-9.98	.95

Variable	k	Missing values	М	SD	Mdn	Mode	Range	Krippen- dorff's α
Prestudy sleep, wake condition	346	477	7.05	1.05	7.10	6.00	4.00-10.00	.96
Stimulus presentation time	711	112	7961.50	31059.58	3000	3000	200–300000	.93
Scope of learned material	815	8	61.54	82.80	36	40	1–900	.99
Study trials	804	19	1.86	2.77	1	1	0.67–34	.74
Immediate testing trials	766	57	1.39	2.02	1	0	0–18	.81
Learning cutoff criterion	173	650	0.72	0.17	0.65	0.60	0.40–1	1
Reward cuing	818	5						1
No reward cuing	805							
With reward cuing	13							
Memory performance based reward	819	4						.67
No reward	817							
With reward	2							
Announcement of delayed test	342	481						.82
No explicit announcement	178							
With explicit announcement	164							
Self-reported test expectation	55	768						.73
No expectation	52							
Expectation	3							
Congruent learning and testing context	813	10						.71

Variable	k	Missing values	Μ	SD	Mdn	Mode	Range	Krippen- dorff's α
Congruent	793			•				
Incongruent	20							
Self-reported active rehearsal	217	606						.70
No active rehearsal	217							
With active rehearsal	0							
Proactive interference at encoding	823	0						1
No introduction	817							
With introduction	6							
Retroactive interference at encoding	823	0						1
No introduction	814							
With introduction	9							
Retroactive interference at delayed testing	823	0						1
No introduction	770							
With introduction	53							
		F	ocal varial	bles charac	teristics			
Emotionality	819	4						.96
Neutral	696							
Negative	77							
Positive	23							

Variable	k	Missing values	М	SD	Mdn	Mode	Range	Krippen- dorff's α
Emotional	9							·
Mixed	14							
Restudy	822	1						.80
No restudy	621							
With restudy	201							
Testing	822	1						.79
No testing	324							
With testing	498							
Feedback	822	1						1
No feedback	622							
With feedback	200							
Encoding instruction	794	29						.88
Intentional	595							
Incidental	199							
Directed forgetting instruction	821	2						1
No instruction	763							
Remember instruction	25							
Forget instruction	33							
Retrieval procedure	823	0						.76
Recognition	374							
Cued recall	345							
Free recall	104							
		Parit	y check va	riables cha	racteristic	S		
Age	823	0	24.10 ^a	12.74 ^a	22.42 ^a	21.90°	3.46-69.85°	.92

Variable	k	Missing values	М	SD	Mdn	Mode	Range	Krippen- dorff's α
Children and adolescents	57		9.68ª	3.52ª	10.10 ^a	10.10 ^a	3.46-16.00 ^a	
Adults	735		23.31 ^a	5.06ª	22.56ª	21.90ª	19.30–59.00°	
Older adults	31		65.70 ^a	2.61ª	65.93°	63.50 ^a	61.27–69.85 ^a	
Definition of dependent variable	823	0						1
Delayed test	691							
Immediate test - delayed test	132							
TST	688	135	5.01	3.02	6.60	7.50	0.35-9.45	.86
SWS	451	372	0.93	0.85	0.60	1.31	0.00-4.60	.88
REM sleep	451	372	0.68	0.69	0.31	0.00	0.00-2.45	.88
		Expl	oratory va	riables char	acteristics	i		
Stimulus material	823	0						.84
Verbal material	372							
Pictures	249							
Mixed and other material	202							
Sleep study design	823	0						1
Natural sleep and wakefulness	383							
Daytime nap	205							
Nighttime nap	44							
TSD with recovery night	85							
TSD without recovery night	40							
SWS deprivation	32							
REM sleep	34							

Variable	k	Missing values	М	SD	Mdn	Mode	Range	Krippen- dorff's α
deprivation								
Retention interval	813	10	17.15	47.76	12.00	12.00	0.67–720.00	.96
Experimental design	823	0						1
Between-subjects design	668							
Within-subject design	155							
		Со	ntrol varia	bles charac	teristics			
Randomization	823	0						.81
Randomized	418							
Not randomized or unknown	405							
Accuracy of effect size computation	823	0						.91
High	581							
Low	242							

Note. Only healthy drug-free human individuals are included in the meta-analysis. Means, standard deviations, medians, modes, and ranges are reported on the sample level. For chronotype categorization the mean score and the cutoff criteria of the respective questionnaire were used. The "sleep tracker" category includes sleep parameter assessments such as actigraphy or pulse monitoring. In case of aggregated to-be-remembered and to-be-forgotten items, we coded directed forgetting instruction as missing. Except for Giganti et al.'s (2014) study (k = 1), the order of sleep versus wakefulness was counterbalanced in all studies with within-subject design. Prestudy sleep, total sleep time (TST), slow-wave sleep (SWS), rapid eye movement (REM) sleep, and retention intervals are reported in hours. The stimulus presentation time is reported in milliseconds. More detailed information about category coding is provided on the Open Science Framework (OSF; https://osf.io/dr6hw/?view_only=c55b3737b6c642cfb170276088930a04). k = 1 number of effect

sizes; missing values = number of effect sizes excluded due to ambiguous or missing information; PSG = polysomnography; TSD = total sleep deprivation.

^aAge in years; based on available information about the mean age of participants included in the analyses of the respective primary studies (k = 464, Krippendorff's $\alpha = .80$).

Table 2Results of the Main Multiple-Moderator Analysis

Variable	k	m	F	t	g	SE	95% CI	df	р
			1	Focal vari	ables				
Emotionality	777	254	1.66					5, 10.71	.231
Neutral	658	249		14.64	.40	.03	[.35, .45]	145.57	<.001
Negative	76	33		8.33	.58	.07	[.44, .72]	36.19	<.001
Positive	23	11		6.95	.45	.07	[.31, .60]	8.98	<.001
Emotional	9	6		3.77	.48	.13	[.15, .80]	4.76	.014
Mixed	11	6		6.10	.62	.10	[.33, .90]	3.91	.004
Restudy	777	254	9.71					2, 88.37	.002
No restudy	593	176		12.72	.38	.03	[.32, .44]	125.59	<.001
With restudy	184	89		12.07	.55	.05	[.46, .64]	81.00	<.001
Testing	777	254	1.32					2, 58.70	.256
No testing	303	98		10.61	.46	.04	[.38, .55]	90.40	<.001
With testing	474	178		12.03	.40	.03	[.33, .46]	94.03	<.001
Feedback	777	254	3.49					2, 91.55	.065
No feedback	587	183		12.57	.39	.03	[.33, .45]	119.45	<.001
With feedback	190	87		8.79	.52	.06	[.40, .64]	99.33	<.001
Encoding instruction	777	254	1.02					2, 49.52	.318
Intentional	581	212		12.42	.40	.03	[.34, .46]	110.38	<.001
Incidental	196	48		6.35	.49	.08	[.34, .65]	50.89	<.001
Directed forgetting instruction	777	254	0.82					3, 10.59	.467
No instruction	719	239		15.50	.43	.03	[.37, .48]	147.43	<.001
Remember instruction	25	10		4.54	.50	.11	[.26, .74]	11.83	.001

Variable	k	m	F	t	g	SE	95% CI	df	р
Forget instruction	33	11		2.31	.30	.13	[.02, .58]	13.16	.038
Retrieval procedure	777	254	1.39					3, 38.35	.261
Recognition	347	98		9.28	.38	.04	[.30, .46]	62.78	<.001
Cued recall	336	144		10.10	.45	.05	[.36, .54]	63.60	<.001
Free recall	94	48		7.62	.49	.06	[.36, .61]	41.23	<.001
				Parity ch	ieck va	riables	5		
Age	777	254	0.53					3, 20.24	.600
Children and adolescents	44	20		3.69	.41	.11	[.18, .65]	24.96	.001
Adults	705	224		16.16	.43	.03	[.38, .48]	145.96	<.001
Older adults	28	10		1.70	.26	.16	[09, .62]	8.98	.124
Definition of dependent variable	777	254	24.87					2, 41.18	<.001
Delayed test	652	231		14.88	.39	.03	[.34, .44]	148.39	<.001
Immediate test - delayed test	125	65		14.06	.58	.04	[.50, .67]	58.47	<.001
			Exp	lorative v	ariable	es			
Stimulus material	777	254	2.71					3, 48.12	.076
Verbal material	345	142		10.94	.46	.04	[.38, .55]	94.31	<.001
Pictures	235	63		7.98	.44	.06	[.33, .55]	39.37	<.001
Mixed and other material	197	85		6.70	.33	.05	[.23, .43]	63.65	<.001
Sleep study design	777	254	2.77					7, 27.80	.031
Natural sleep and wakefulness	365	138		14.28	.54	.04	[.46, .61]	115.89	<.001
Daytime nap	181	60		6.01	.34	.06	[.22, .45]	37.64	<.001
Nighttime nap	44	10		2.62	.54	.21	[.08, 1.01]	10.18	.025
TSD with	85	23		1.14	.14	.13	[13, .42]	11.35	.278

Variable	k	m	F	t	g	SE	95% CI	df	р
recovery night			-						
TSD without recovery night	36	17		3.25	.30	.09	[.10, .49]	15.00	.005
SWS deprivation	32	9		2.33	.31	.13	[.02, .59]	12.04	.038
REM sleep deprivation	34	10		6.45	.43	.07	[.28, .58]	11.87	<.001
Experimental design	777	254	1.87					2, 76.54	.176
Between- subjects design	633	203		14.05	.44	.03	[.38, .50]	128.68	<.001
Within-subject design	144	51		6.03	.34	.06	[.23, .46]	58.12	<.001
			C	ontrol vai	riables				
Randomization	777	254	0.01					2, 103.92	.912
Randomized	392	130		10.39	.42	.04	[.34, .50]	95.35	<.001
Not randomized or unknown	385	126		10.59	.43	.04	[.35, .51]	102.82	<.001
Accuracy of effect size computation	777	254	0.79					2, 78.35	.375
High	546	190		14.10	.41	.03	[.35, .47]	132.98	<.001
Low	231	94		10.25	.45	.04	[.37, .54]	82.07	<.001

Note. Heterogeneity statistics of the intercept-only three-level model without RVE: Q(822) = 1878.56, p < .001; $\tau^2 = 0.15$; $\sigma_b^2 = 0.12$; $\sigma_w^2 = 0.02$; $I_t^2 = 60.00\%$; $I_b^2 = 50.30\%$; $I_w^2 = 9.70\%$; ICC = .84. Results for the continuous variable retention interval: $\beta = -0.001$, SE = 0.002, 95% CI [-0.01, 0.01], t(2.11) = 0.37, p = .744, k = 777, m = 254. k = number of effect sizes; <math>m = number of independent samples; $F = approximate Hotelling's <math>T^2$ [AHT] F value; t = small-sample corrected t value; t = small-sample corrected t = small-sample cor

effect size heterogeneity; τ^2 = total variance; σ_b^2 = variance between samples; σ_w^2 = variance within samples; I_t^2 = total amount of heterogeneity relative to the total amount of variance; I_b^2 = amount of heterogeneity between samples relative to the total amount of variance; I_w^2 = amount of heterogeneity within samples relative to the total amount of variance; ICC = intraclass correlation coefficient, quantifies effect size correlation within samples.

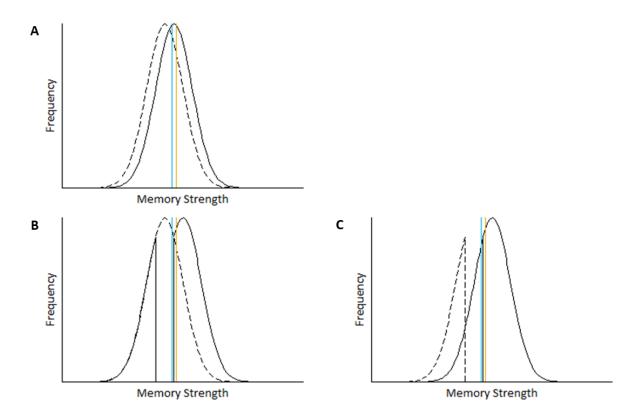
Table 3Results of the Three-Parameter Selection Model (3PSM) and Likelihood Ratio Test

Subset	k	τ^2	β	ψ	χ²(1)	р
1	42	0.09	0.30	0.52	1.14	.285
2	42	0.09	0.35	0.62	0.58	.445
3	42	0.10	0.44	0.63	0.81	.368
4	41	0.03	0.30	0.63	0.44	.507
5	41	0.19	0.68	2.40	2.44	.118
6	41	0.47	0.79	4.05	5.50	.019
7	41	0.12	0.49	2.15	1.48	.224
8	41	0.17	0.45	1.34	0.24	.624
9	41	0.12	0.30	0.69	0.28	.600
10	41	0.11	0.40	1.17	0.05	.823
11	41	0.14	0.42	0.64	0.61	.435
12	41	0.13	0.47	1.64	0.73	.393
13	41	0.11	0.44	1.12	0.04	.848
14	41	0.08	0.05	0.13	8.51	.004
15	41	0.03	0.23	0.26	4.01	.045
16	41	0.11	0.47	1.70	0.68	.408
17	41	0.19	0.50	0.91	0.03	.867
18	41	0.16	0.36	0.83	0.11	.745
19	41	0.07	0.39	0.77	0.21	.649
20	41	0.02	0.25	0.46	1.22	.269

Note. k = number of effect sizes; τ^2 = heterogeneity between effect sizes; β = corrected overall effect size estimate; ψ = probability ratio of observing a published nonsignificant result relative to a significant one.

Figure 1

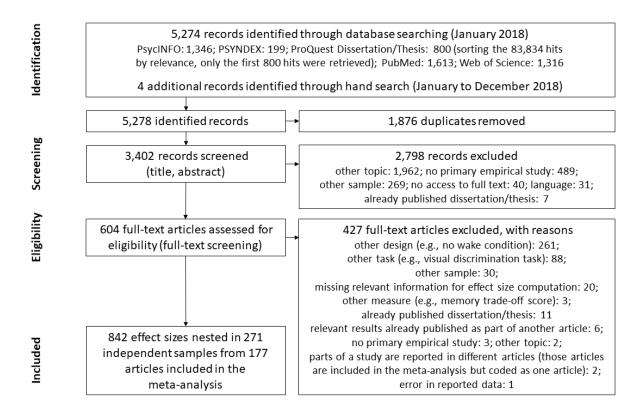
Illustration of Predictions for Moderators "Restudy", "Testing", "Feedback", and "Encoding Instruction" Based on Interference Theories



Note. Dashed curves represent memory strength in no-restudy, no-testing, no-feedback, and incidental-learning conditions (i.e., weak encoding conditions). Solid curves represent memory strength in restudy, testing, feedback, and intentional-learning conditions (i.e., strong encoding conditions). Blue and yellow vertical lines represent retrieval thresholds for sleep and wake conditions, respectively. Retrieval thresholds are lower for sleep than for wake conditions because, according to interference theories, sleep reduces retroactive interference, thereby making memories more accessible and easier to retrieve. The area under the curves on the right side of blue or yellow lines represent percentages of retrieved items in a memory test after the retention interval. The area under the curves between blue and yellow lines represent the sleep benefit. Panel A: Predictions of interference theories for moderators "restudy" and "encoding instruction". As is apparent when comparing the areas under the dashed and solid curves between the two retrieval

encoding conditions. Panel B: Prediction of interference theories for moderator "testing". In testing conditions, retrieval practice during learning bifurcates memory strength distributions because successfully retrieved items gain memory strength whereas not retrieved or not tested items remain on the same level of memory strength. Due to this bifurcation, the sleep benefit in memory retrieval is almost equal for testing and no-testing conditions as can be seen by comparing the areas under the dashed and solid curves between the two thresholds. Panel C: Predictions of interference theories for moderator "feedback". Whereas memory strength distributions of successfully tested items (to the right of the solid vertical line) are identical for feedback and no-feedback conditions, they differ markedly for not successfully tested items (to the left of the solid vertical line). As is apparent when comparing the areas under the dashed and solid curves between the two retrieval thresholds, the sleep benefit is larger if feedback is provided after testing than if it is not. The figures were derived from real data reported by Denis et al. (2020). For details of the derivation see the supplemental materials.

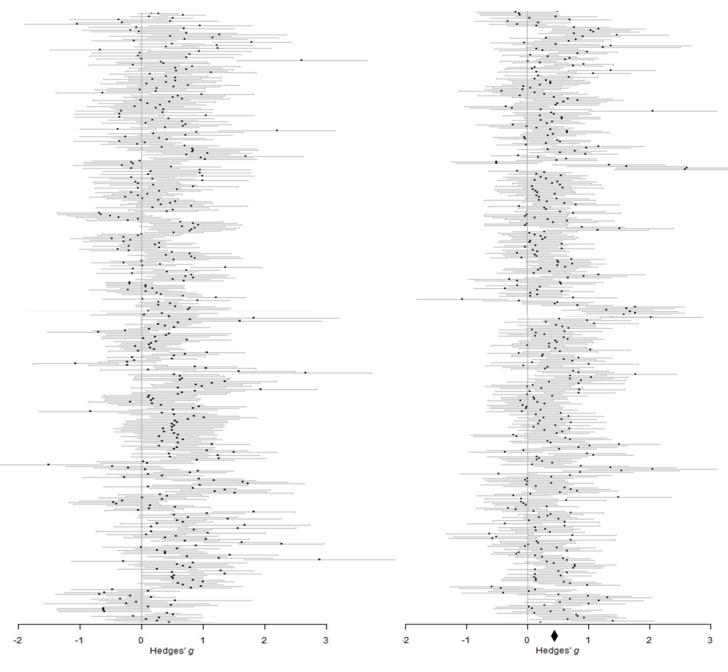
Flow Diagram of the Study Selection Procedure



Note. Flow diagram based on the PRISMA Statement by Moher et al. (2009). The flow diagram shows the study selection procedure. Note that the final data set of 842 effect sizes includes 823 effect sizes for episodic memory measures and 19 effect sizes for response bias *c*. As the focus of the current meta-analysis is on episodic memory, we excluded the 19 effect sizes from the analysis.

Figure 3

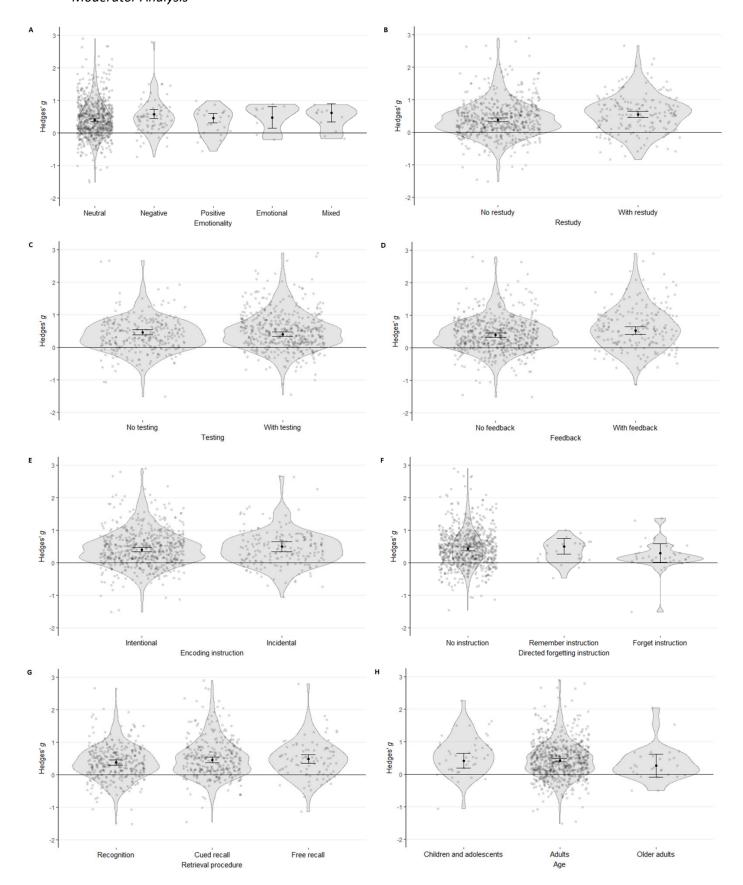
Forest Plot for the Overall Effect of Sleep on Episodic Memory

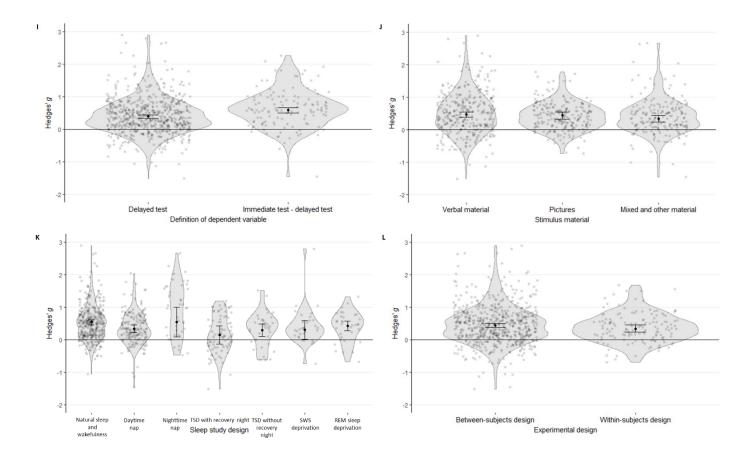


Note. The forest plot shows effect size estimates and their confidence intervals in the order of the publication year. For a more detailed forest plot see Figure S1 in the supplemental materials. Table S3 provides the exact numerical values for the depicted effect sizes and confidence intervals. The overall effect of sleep on episodic memory (i.e., g = 0.44) was estimated using an intercept-only three-level model with RVE.

Figure 4

Effect Sizes for Categorical Focal, Parity Check, and Exploratory Variables in the Main MultipleModerator Analysis

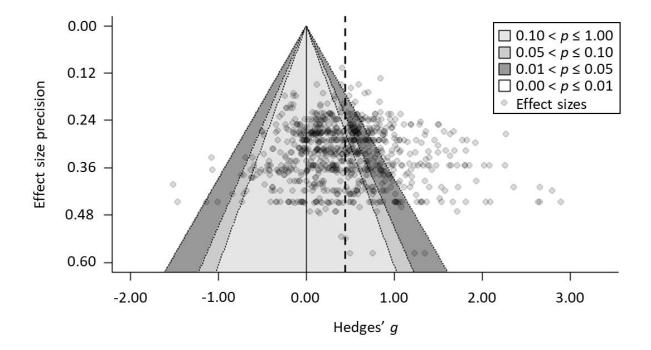




Note. Solid black dots represent the marginal average effects after adjusting for variation in the full set of potential moderators included in the moderator analysis reported in Table 2. Gray dots represent effect sizes from primary studies. Error bars show 95% confidence intervals. TSD = total sleep deprivation; SWS = slow-wave sleep; REM = rapid eye movement.

Figure 5

Contour-Enhanced Funnel Plot for the Overall Effect of Sleep on Episodic Memory



Note. To measure effect size precision, a function of the sample size is used. This modified measure of effect size precision was calculated as $[(n_{sleep} + n_{wake})/(n_{sleep} * n_{wake})]^{1/2}$ for between-subjects designs and $1/N^{1/2}$ for within-subject designs (Pustejovsky & Rodgers, 2019).