

# Sleep on it? Re-examining the Impact of Sleep on Memory

David Philip Morgan

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in  
the Department of Psychology.

Royal Holloway, University of London

2020



## Declaration of Authorship for Co-Authored Work

If you are presenting partly co-authored work, please indicate below your individual contribution to the thesis.

Name of candidate: David Philip Morgan

Thesis title: Sleep on it? Re-examining the impact of sleep on memory

I confirm that the thesis that I am presenting has been co-authored with:

Dr Jakke Tamminen, Professor Laura Mickes, Dr Travis Morgan Seale-Carlisle

Within this partly co-authored work, I declare that the following contributions are entirely my own work:

Chapter 1, General Introduction: Entirely my own work

Chapter 2, Examining the Impact of Sleep on Eyewitness Identifications:

I designed the experiment, collected the data (Dr Seale-Carlisle assisted with data collection), analysed the data, and wrote the first draft of the manuscript for publication. Dr Tamminen and Professor Mickes provided critical comments on revisions.

Chapter 3, Norms for Sleepiness Scales and Sleep Duration in a Young Adult American Sample:

I designed the experiment, collected the data (Dr Seale-Carlisle assisted with data collection), analysed the data, and wrote the first draft of the manuscript for publication. Dr Tamminen and Professor Mickes provided critical comments on revisions.

Chapter 4, No Effect of Sleep on the Misinformation Effect:

I designed the experiment, collected the data, analysed the data, and wrote the first draft of the manuscript for publication. Dr Tamminen and Professor Mickes provided critical comments on revisions.

Chapter 5, Re-Examining the Impact of Sleep on Emotional Memory:

I designed the experiment, collected the data, analysed the data, and wrote the first draft of the manuscript for publication. Dr Tamminen and Professor Mickes provided critical comments on revisions.

Chapter 6, General Discussion: Entirely my own work.

Signed: ..... Date: 31<sup>st</sup> October 2019  
(Candidate)

Signed: ..... Date: 31<sup>st</sup> October 2019 (Supervisor)

## Acknowledgements

There are many people that I owe a huge debt of gratitude to. The first is Professor Laura Mickes, who took me under her wing in 2014. Since, she has helped me to develop and refine my skills as a scientist with infectious enthusiasm and has taught me so so so much of the years since I joined her lab. I can honestly say that Laura is an admirable and inspirational scientist who at every turn has supported me both academically and emotionally. I will certainly miss the laughter that came from her office when we worked together, honestly some of the best years of my life so far. I am eternally grateful to Laura, she really has changed my life and my families' lives forever, and I can certainly say that I made not just a colleague but friend for life.

Likewise, Dr Jakke Tamminen played a major role in shaping the scientist I am today. I am very lucky to have had his mentorship. Jakke simultaneously challenged me intellectually to develop new innovative ideas whilst also allowing me to pursue replication efforts of past research. Jakke always supported me via friendly intellectual debate, which helped me clarify and adjust my approach and understanding on strategies to achieve scientific progress. I am very grateful to Jakke for taking me under his wing for the past three years and for always acting with my best interests at heart.

It goes without saying that this experience would not have been the same without the jest and audible laughter emanating from the PhD office. We all shared the peaks and much longer troughs associated with completing a PhD, but despite that I could always count on the exceptional quality of banter in the PhD office to cheer me up. It is understood that my cohort had the best banter (other cohorts may not agree), so I am sending a massive thank you to the PhD class of 2016 (Adam, Jasmine, Rachael, Franziska, Kathrin, Craig, Maria, Hanna and Lilla). I'd like to give a special shout out to the following: Rachael L, Devin, Michaela, Vikki, Gita, Rebecca, Rachel N, Jen, Beatrice, Ben, Isaac, Jasmine, Adam, Lore, Travis, Chloe and

Becky. Thank you for making this experience so incredibly enjoyable, it would not have been the same without you.

Lastly I have to thank my family. My mother's diagnosis is something that has undoubtedly had a profound impact on all of our lives. To my mother, Judith, although we didn't get as much opportunity as I would like to talk, you have always shown love and compassion towards me and I could always count on you for a hug when things get tough. To my father, Glyn, the bravest man I know, thank you for always supporting and guiding me to pursue what I want to do, not what others dictate to me (within reason of course). To my sister, Sarah, thank you for encouraging me to get my butt to school and pursue my dreams. As young carers, the world did not promise much to myself and my sister, we had to work harder than most to get to where we are today, and this came with its sacrifices. Importantly, this is not a testament to the notion that "working hard enough will make you successful". Hard work is one side of the coin, opportunity is also necessary. I was fortunate enough to be presented with the opportunity to extend my career. Many young carers are not presented with the opportunities that I was afforded and it will be one of my aims as part of the path that this PhD has brought me on to extend those opportunities to others like me and my sister who experienced the harsh realities of mental and physical health from a very young age.

## Abstract

Sleep is widely believed to play an active role in memory consolidation – the process by which memories are stored. Many studies have demonstrated that sleep yields greater memory performance on various tasks in comparison to an equivalent period of wake. However, the studies conducted in this thesis reveal a divergent pattern of results which do not consistently demonstrate the benefits of sleep on memory. The first study, a registered report and the largest sleep vs. wakefulness comparison to date ( $N = 4,000$ ), examined the impact of sleep on eyewitness identification performance. It was predicted that sleep compared to wake would benefit accuracy. The data did not support that prediction: there were no differences between sleep and wake. The second study used the questionnaire data collected as part of the registered report ( $N = 7,533$ ) to provide normative data of sleepiness scales and sleep duration for clinicians and researchers alike to use. The third study, a conceptual replication and extension of previous research, investigated the impact of sleep on integrating true and false information which would reduce memory accuracy. Sleep was not detrimental to memory accuracy, nor was there strong evidence that it was beneficial. Given the absence of the purported effects of sleep in these experiments, a replication experiment was designed to re-examine well-known findings of sleep on memory. We attempted to replicate research indicating that sleep preferentially benefits the consolidation of emotional stimuli. No evidence was found in support of these findings, but there was a sleep benefit on overall memory (however there was also a potential time of day confound). Potential explanations for the mixed findings are considered and large-scale replication efforts are proposed to clarify understanding of the benefits of sleep on memory.

## Table of Contents

<b>List of Tables and Figures .....</b>	<b>8</b>
<b>Bibliography of Work Completed During the PhD .....</b>	<b>10</b>
<b>Chapter 1: General Introduction .....</b>	<b>11</b>
Sleep and Memory Consolidation .....	11
Meta-Methods .....	20
Summary of the current thesis.....	25
<b>Chapter 2: Examining the Impact of Sleep on Eyewitness Identifications.....</b>	<b>33</b>
Abstract .....	33
Introduction .....	34
Methods.....	39
Analysis Strategy.....	44
Results .....	47
Discussion .....	53
<b>Chapter 3: Norms for Sleepiness Scales and Sleep Duration in a Young Adult American Sample.....</b>	<b>62</b>
Abstract .....	62
Introduction .....	64
Methods.....	68
Results .....	71
Discussion .....	78
<b>Chapter 4: No Effect of Sleep on the Misinformation Effect.....</b>	<b>85</b>
Abstract .....	85
Introduction .....	86
Methods.....	93
Analysis Strategy.....	98

Results .....	100
Discussion .....	107
<b>Chapter 5: Re-Examining the Impact of Sleep on Emotional Memory .....</b>	<b>113</b>
Abstract .....	113
Introduction .....	114
Methods .....	121
Analysis Strategy.....	125
Results .....	127
Discussion .....	135
<b>Chapter 6: General Discussion .....</b>	<b>141</b>
Summary of findings .....	142
Critical Analysis of Findings.....	151
Limitations .....	158
Future Directions.....	163
Conclusion.....	166
Appendices.....	168
Appendix A .....	168
References.....	170

## List of Tables and Figures

<b>Table 1.</b> Demographic information by sleep, wake, AM control and PM control groups.....	47
<b>Table 2.</b> Frequencies of correct IDs (CID), filler IDs (FID), and no-IDs for sleep and wake groups.....	48
<b>Figure 1</b> ROC curves and CAC plots for sleep and wake groups.....	49
<b>Table 3.</b> Frequencies of correct IDs (CID), filler IDs (FID), and no-IDs for AM and PM control groups. ....	51
<b>Figure 2.</b> ROC curves and CAC plots for AM and PM control groups.....	52
<b>Figure 3.</b> ROC curves and CAC plots for experimental and control groups. ....	60
<b>Table 4.</b> Demographic data for all participants. ....	69
<b>Table 5.</b> Demographic data of the subset who completed the SSS and ESS. ....	70
<b>Figure 4.</b> Average SSS scores by sex, time of day and education. ....	73
<b>Table 6.</b> SSS scores split by time of day, education, gender, and age category. ....	74
<b>Figure 5.</b> Average ESS scores by sex, time of day and education.....	76
<b>Table 7.</b> ESS scores split by time of day, education, gender, and age category. ....	77
<b>Figure 6.</b> Hours of sleep by age (collapsed across sex and education) in the current study (Morgan et al.) and Maslowsky and Ozer (2014).. ....	79
<b>Table 8.</b> Sleep duration scores split by time of day, education, gender, and age category. ....	80
<b>Figure 7.</b> Procedure for the experimental conditions.....	97
<b>Table 9.</b> Demographic characteristics of participants per condition. ....	102
<b>Table 10.</b> Means, standard deviations and 95% confidence intervals for each type of memory per experimental condition. ....	103
<b>Table 11.</b> Means, standard deviations and 95% confidence intervals for each type of memory per control condition. ....	104



<b>Table 12.</b> Means, standard deviations and 95% confidence intervals for each type of memory.	106
<b>Table 13.</b> Means, standard deviations and 95% confidence intervals for each type of memory per experimental and control conditions.	108
<b>Table 14.</b> Demographic characteristics per condition.	128
<b>Figure 8.</b> Mean $d'$ across negative and neutral items for sleep and wake conditions and AM and PM conditions	129
<b>Table 15.</b> Mean $d'$ , standard deviation, and 95% confidence intervals for negative and neutral stimuli per experimental condition.	130
<b>Figure 9.</b> Calibration plots for neutral and negative images for experimental and control conditions	134
<b>Table 16.</b> Summary of findings in the thesis	142
<b>Figure 10.</b> Simulations of Clemens et al. (2006) data with a small ( $N = 15$ ) and large sample size ( $N = 1000$ ).	162

## Bibliography of Work Completed During the PhD<sup>1</sup>

Morgan, D. P., Tamminen, J., Seale-Carlisle, T. M., & Mickes, L. (2019). Investigating the impact of sleep on eyewitness identifications. *Royal Society: Open Science*, 6, 170501.

Morgan, D. P., Tamminen, J., & Mickes, L. (In preparation). Norms for sleepiness scales and sleep duration in a young adult American sample.

Morgan, D. P., Mickes, L., & Tamminen, J. (In preparation). No Effect of Sleep on the Misinformation Effect.

Morgan, D. P., Tamminen, J., & Mickes, L. (In preparation). Re-Examining the Impact of Sleep on Emotional Memory.

---

<sup>1</sup> This thesis has been submitted in the alternative format. Papers being submitted: Three are in preparation for publication and one has been accepted for publication.

## Chapter 1: General Introduction

### Sleep and Memory Consolidation

In the most cited review on sleep and memory literature to date Diekelmann and Born (2010) stated that

“Numerous studies have confirmed the beneficial effect of sleep on declarative and procedural memory in various tasks, with practically no evidence for the opposite effect (sleep promoting forgetting). Compared with a wake interval of equal length, a period of post-learning sleep enhances retention of declarative information and improves performance in procedural skills.” (p.1).

This statement represents a commonly shared viewpoint amongst researchers who study the relationship between sleep and memory. However, the pattern of results from the research presented in this thesis is mixed when compared to this statement. Below I provide a meta-context of the research which motivated the experiments presented in this thesis.

Consolidation refers to the process by which new memories, experiences, or new information are transferred into long-term memory, a mechanism first proposed by researchers between 1880 and 1910 (e.g., Burnham, 1903; McDougal, 1901, Muller & Pilzecker, 1900). This proposal was based on findings that if new memories are not given time to stabilise, they can become susceptible to interference from incoming information. The role of sleep in this process was considered in response to finding that the forgetting of new information did not decrease linearly over time, but exponentially with some information being retained over time (Ebbinghaus, 1885). Given that there was a period of sleep between study and test in Ebbinghaus’s experiment, Jenkins and Dallenbach (1924) proposed that sleep may be involved in the retention of newly learnt information in Ebbinghaus’s (1885) experiment.

Jenkins and Dallenbach (1924) asked two participants to learn nonsense syllables either in the morning or the evening. They found that when participants slept between study and testing they had greater recall compared to a period of wake. This finding was later conceptually replicated about a decade later (Van Ormer, 1932). This idea was not revisited until the 1960's, where researchers continued to find that sleep compared to a period of wake, during the day (i.e., not sleep deprived), yielded greater memory (e.g., Barret & Ekstrand, 1972).

Over time, researchers found that sleep can benefit consolidation of both motor and declarative forms of memory (see Diekelmann & Born, 2010). To investigate this, participants learn a motor task or list of items (e.g., words, images, etc.) and are tested on their memory after a period of sleep or wake. For instance, sleep has been found to benefit performance on finger tapping tasks (e.g., Brawn, Fenn, Nusbaum & Margoliash, 2010; Fischer et al., 2005; Korman et al., 2003, 2007; Walker, Brakenfield, Morgan, Hobson & Stickgold, 2002; Walker, Stickgold, Alsop, Gaab & Schlaug, 2005; but see Rickard et al., 2008; Cai & Rickard, 2009), free recall (e.g., Schönauer, Pawliziki, Köck & Gais, 2014), associative memory (e.g., Drosopoulos, Schulze, Fischer & Born, 2007; Payne et al., 2012; Tucker & Fishbein, 2008; Wilhelm et al., 2011; Tucker, Tang, Uzoh, Morgan & Stickgold, 2011), and recognition memory (e.g., Drosopoulos, Wagner & Born, 2005; Wagner, Kashyap, Diekelmann & Born, 2007).

#### *Passive or Active Memory Consolidation*

*Active Memory Consolidation.* Broadly there are two overarching theoretical accounts of the finding that sleep benefits memory (Ellenbogen, Payne & Stickgold, 2006). The first posits that sleep plays an active role in memory consolidation (Diekelmann & Born, 2010). This is because sleep specific brain activity that only occurs during sleep, such as slow wave

sleep (SWS; e.g., Born, 2010; Walker, 2009), sleep spindles (e.g., Clemens, Fabo & Hålasz, 2005), and hippocampal ripples (e.g., Norman et al., 2019; Rosanova & Ulrich, 2005, Fernández-Ruiz et al., 2019) have all been positively associated with greater memory performance (but see Ackermann et al., 2015; Mantua, 2018; Pan & Rickard, 2015). In fact, numerous reviews propose that sleep benefits memory (e.g., Diekelmann & Born, 2010; Feld & Diekelmann, 2015; Diekelmann, Wilhelm & Born, 2009; Stickgold, 2005; Walker, 2009; Walker & Stickgold, 2004). This conclusion has permeated the literature. The two following quotes are such representative examples:

“there is now compelling evidence that the long-term storage of memories preferentially occurs during sleep” (p. 1563, Wilhelm et al., 2011);

“it is well established that sleep contributes to memory consolidation processes”, (p.1, Igloi, Gaggioni, Sterpenich & Schwartz, 2015).

One hypothesis about the active role of sleep in memory consolidation is the Systems-Level Consolidation Hypothesis (see review Born & Wilhelm, 2012). This hypothesis suggests that there are two primary memory stores, the hippocampus and the Neocortex (Marr 1971; McClelland, McNaughton, & O'Reilly 1995). The hippocampus rapidly encodes and temporarily stores newly acquired information (Diekelmann & Born, 2010). When new information is encoded, a distributed network of brain regions stores specific information related to that new memory (e.g. faces, locations, etc.), but the neuronal representation of that information is first primarily stored in the hippocampus (Diekelmann & Born, 2010). During subsequent sleep, specifically SWS, hippocampal neurons repeatedly reactivate this new information.

Of course, the Systems Level Consolidation Hypothesis does not propose that the hippocampus operates in isolation, but actively communicates with the Neocortex during sleep

to ensure long-term storage of recently encoded information. This aforementioned reactivation allows memory representations to be redistributed throughout the Neocortex, where new memories are gradually integrated into pre-existing memories (Diekelmann & Born, 2010). This process of redistributing newly acquired information can be a slow process occurring over months and years, and so the Neocortex is regarded as slow-learning and is referred to as the long term memory store. Specifically, this redistribution of newly acquired information into the Neocortex relies heavily on reactivation of neurons during SWS where slow oscillations (<1 Hz; originating in the Neocortex), ripples (~80 Hz; originating in the hippocampus) and spindles (~12–15 Hz; originating in the thalamus; Diekelmann & Born, 2010) all contribute to actively transferring new information from the hippocampus into the Neocortex.

The reactivation of memories in the hippocampus and Neocortex has been demonstrated in animal models and more recently in humans. For instance reactivation of spatial memories in mice has been identified during sleep (and during wakeful rest; Davidson, Kloosterman & Wilson, 2009), whereby place cells which fire whilst encoding a new environment replay within the hippocampus in a similar pattern (e.g., Wilson & McNaughton, 1994). Recently this reactivation has been demonstrated to contain information about visual stimuli learned during wakefulness in humans (Schönauer et al., 2017). Schönauer et al. (2017) found that classifiers trained on sleep recording data can classify which visual stimuli have been learned before a period of sleep, suggesting that the neurons which represent these stimuli are reactivated during sleep. Importantly, the classification accuracy of activity recorded during SWS was positively correlated with greater memory performance after sleep on a memory task. In other words, if the classifier using SWS data was more certain that a participant witnessed a particular set of stimuli then this certainty was related to the participant's higher accuracy presumably because the studied information was reactivated. Indeed it should be expected that activity observed during SWS in humans should contain

information related to recently acquired memories if that information was reactivated as the Systems Level Consolidation Hypothesis proposes.

Further evidence indicates that during SWS a temporal contingent looping cascade of activity involving the upstate and downstate of slow oscillations nested ripples and spindles ultimately contribute to systems consolidation (for a review see Klinzing, Niethard & Born, 2019). For example, researchers have proposed that the upstate of slow oscillations triggers spindles which are nested within the upstate of an oscillation (Klinzing, Niethard & Born, 2019). These spindles communicate with the hippocampus to coordinate ripples and in turn reactivation of the neurons which represent a recently encoded memory (Diekelmann & Born, 2010; Klinzing, Niethard & Born, 2019). Spindles have also been found to reach into the neo cortex which is believed to simultaneously facilitate consolidation and synaptic plasticity within networks associated with the memory being replayed, therefore actively enabling information to be transferred between the hippocampus and the neo cortex (Diekelmann & Born, 2010; Klinzing, Niethard & Born, 2019).

It is precisely this process that explains the findings that sleep yields greater memory performance when compared to a period of wake where this process does not occur (e.g., Brawn, Fenn, Nusbaum & Margoliash, 2010; Drosopoulos, Schulze, Fischer & Born, 2007, Drosopoulos, Wagner & Born, 2005). Additionally researchers propose that reactivation integrates new memories with pre-existing networks (e.g., Tamminen, Lambon Ralph & Lewis, 2013; Tamminen, Payne, Stickgold, Wamsley & Gaskell, 2010). Reactivation can also explain how memories become less reliant on the hippocampus over time (e.g., Sirota Csicsvari, Buhl & Buyaski, 2003; Staresina et al., 2015) and in turn more resistant to interference (e.g., Diekelmann, Büchel, Born, Rasch, 2011).

Although the systems consolidation hypothesis is a popular theory which can explain many findings related to consolidation of new memories and their integration into existing networks, it does not describe the cellular processes which are also involved in consolidation. One theory that does describe this process is the Synaptic Homeostasis Hypothesis (Tononi & Cirelli, 2003; Tononi & Cirelli, 2014). Broadly this hypothesis proposes that the purpose of sleep is to downscale the synaptic strength of neurons in the brain so that the system does not become overwhelmed, and allows for sustainability relative to the size of the organism and the energy which it demands (Tononi & Cirelli, 2003; Tononi & Cirelli, 2014). The first central tenet of this theory is that when information is encoded synaptic strength is increased within neurons which represent the encoded information. This process of synaptic strengthening is known as long-term potentiation, LTP, a cascade of synaptic activity which induces synaptic plasticity in the hippocampus and the neo cortex enabling encoding and stabilisation of encoded information (Silva, 2003). The induction of LTP can occur when new information is encoded which can lead to an influx of glutamate, a neurotransmitter which is crucial for learning and memory (Riedel, Platt & Micheau, 2003). This allows  $\text{Ca}^{2+}$  to pass through NMDA post-synaptic receptors which contributes to synaptic plasticity and eventually LTP within neurons which represent the memory at hand (Mednick, Cai, Shuman & Anagnostaras, 2011).

Another tenet of the synaptic homeostasis hypothesis is that LTP occurring during wake should be directly related to the slow wave activity occurring during sleep in order to ensure homeostasis (Tononi & Cirelli, 2003; Tononi & Cirelli, 2014). For instance, if a greater induction of LTP occurs during waking periods, then there should also be greater slow wave activity during sleep and that this activity should be localised to areas in the neo cortex related to the information that induced the LTP. Critically, in this sense the primary role of sleep in this sense is to downscale the synaptic strength (i.e. potentiation) of neurons which has occurred during waking periods to a sustainable magnitude so that further LTP can occur



during subsequent waking periods (Tononi & Cirelli, 2003; Tononi & Cirelli, 2014). Indirect evidence for this theory has shown that the amount of SWS is higher during early periods of sleep when potentiation is high, and decreases throughout the duration of sleep where potentiation decreases, leaving the synapses de-potentiated during REM sleep which is more prominent at later stages of sleep (Niethard & Born, 2019). Furthermore, SWS is also found to reduce synaptic plasticity and in turn subsequent LTP, which may presumably allow for downscaling to occur and prevents interference (Mednick et al., 2011). Finally Tononi & Cirelli (2006) propose that sleep strengthens memory by reducing signal to noise ratios at the synapses via downscaling whereby memories represented by strongly potentiated synapses are preserved and memories represented by weakly potentiated synapses are forgotten (Diekelmann & Born, 2010). This homeostatic process is thought to drive the cellular consolidation of memories during sleep and lead to greater memory performance and better encoding of new memories following a period of sleep.

It should be noted however that evidence on support of this hypothesis is mixed. Indeed, it does appear that slow oscillations do result in a global reduction of potentiation in neurons, however other evidence indicates that if learning has occurred prior to sleep this can actually increase potentiation in certain regions of the brain after sleep (Diekelmann & Born, 2010). Additionally this hypothesis also presupposes that weak memories are simply forgotten which is not necessarily the case (Diekelmann & Born, 2010; Mednick et al., 2011). For example, it implies that if a memory is weak then it is unlikely to withstand the downscaling process relative to other memories which are more strongly potentiated. This may not necessarily be the case since it has been demonstrated that perhaps memories which could be weak and are associated with reward or emotion can be preserved following a period of sleep (e.g. Drosopoulos, Schulze & Born, 2007; Hu et al., 2008; Igloi, Gaggioni, Sterpenich & Schwart, 2015). Nevertheless, other additive active theories of sleep-dependent memory consolidation

can account for this issue, such as the Synaptic Tag and Capture Hypothesis, which allows for weakly potentiated synapses to be stabilised via co-activation of the hippocampus and relevant brain regions (e.g., emotional memory and the amygdala; see Frey & Morris, 1997). Regardless, synaptic downscaling is generally regarded to contribute to memory consolidation by reducing signal to noise ratios of memories during SWS thus facilitating long term storage of those memories and in turn allowing new encoding to occur (Diekelmann & Born, 2010).

*Passive Memory Consolidation.* Thus far I have discussed theories pertaining to an active role of sleep in memory consolidation, which is the prevailing consensus amongst most sleep scientists, but some propose that the role of sleep is not active and is instead passive (Ellenbogen, Payne & Stickgold, 2006). For instance, the Passive Interference Reduction Hypothesis suggests that sleep does not contribute to consolidation via reactivation, downscaling or by providing optimal conditions for consolidation to occur, but simply shelters encoded information from retroactive interference (i.e. memory for recently studied information is disrupted by encoding new information shortly afterwards; Jenkins & Dallenbach, 1924). Here, it is predicted that sleep will reduce forgetting because it leads to similar brain state to anterograde amnesia where new information cannot be encoded. Critically this hypothesis does not account for a consolidation process. However, studies using interference paradigms are thought to make this argument unsustainable, where researchers have demonstrated that sleep also makes information encoded before a night's sleep more resistant to interference (Ellenbogen, Hulbert, Stickgold, Dinges, Thompson-Schill, 2006). If sleep simply protected memories from interference then any learning which occurred after sleep would be expected to disrupt those memories, this was not the case. Therefore researchers have rejected the notion that sleep simply passively protects memories from interference (Ellenbogen, Payne & Stickgold, 2006).

But, it should be noted that those studies which demonstrated that sleep protects memories from interference recently failed to replicate (Bailes, Cadwell, Wamsley & Tucker, 2020). Moreover, computational models can explain changes in memory performance over time without accounting for a period of consolidation such as that during sleep (e.g., Ecker, Brown & Lewandowsky, 2014). However, given the numerous studies which demonstrate an active role of sleep in memory consolidation, it seems that further research will be required to delineate the exact role sleep plays here. Researchers are yet to reconcile the findings of sleep research with the computational models which explain data better without consolidation. Regardless, it is plausible that reduced interference does play a role in aiding memory consolidation to occur and that sleep simultaneously has an active role in aiding memory consolidation whilst also protecting memories from interference (Mednick, et al., 2011).

As is reported below in our Meta-Methods I did not collect physiological data or conduct and so cannot directly relate behavioural performance on our memory tasks to sleep specific activity. This means that for one of my experiments, presented in Chapter 2, any benefits of sleep or lack thereof that are found could not be related to either the passive or systems consolidation hypotheses uniquely. For example, if an overall benefit of sleep emerged for memory compared to a period of wake it is unclear whether sleep passively protected the memory from interference because the relationship between memory performance physiological activity were not measured. Thus benefits of sleep cannot be attributed to an active account of sleep. Nevertheless in two other experiments presented in Chapters 4 and 5 I was able to conduct manipulations which may indirectly speak to the systems consolidation hypothesis or the synaptic homeostasis hypothesis to test for an active role of sleep in memory consolidation. For example, in the experiment presented in Chapter 5 participants studied negative and neutral images and either slept or remained awake and were tested on their memory for those images to determine whether sleep selectively consolidates emotional

memories over and above neutral memories when compared to wake. If sleep was found to selectively consolidate emotional memories when compared to wake, this would suggest that sleep has an active involvement in memory consolidation because this goes above and beyond protection from retroactive interference proposed by passive accounts of sleep and memory consolidation. Broadly the general idea that sleep benefits memory and, more specifically the Systems-Level Consolidation Hypothesis and Synaptic Homeostasis hypothesis motivated the experiments conducted and presented in this thesis. However the manipulations which were set up to test those hypotheses did not reveal evidence for an active role of sleep in memory consolidation, and where sleep benefits are found they cannot be exclusively related to active or passive accounts of sleep dependent memory consolidation. In fact the current state of the literature and the results reported here, suggest that our understanding of the impact of sleep on memory may need to be re-examined.

## Meta-Methods

### *AM-PM: PM-AM Design*

In a standard memory experiment, participants are given some new information to learn or have to learn a new task, known as the study phase. Participants are later tested on their memory for the information that they had learnt during the study phase, known as the test phase. In sleep research, this standard design is integrated with an AM-PM: PM-AM design to test the impact of sleep on memory. In the AM-PM: PM-AM design participants are assigned to a sleep or a wake condition. In the sleep condition participants complete the study phase in the evening (PM) and complete the test phase in the morning (AM). Comparatively, participants assigned to a wake condition complete the study phase in the morning (AM) and complete the test phase in the evening (PM). The wake condition is used as a comparison because neural activity occurring during sleep does not occur during wake. This design allows

researchers to experimentally manipulate sleep and wake between the study and test phases to evaluate the impact of sleep on memory.

An obvious limitation of this design is the possibility of time of day effects (e.g., Schmidt, Collette, Cajochen & Peigneux, 2007). Take a scenario where participants in a sleep condition had better memory performance than participants in the wake condition. One conclusion might be that sleep-based consolidation benefitted memory and therefore led to greater performance on the task. Another possibility, however, is that memory in the sleep condition was better because participants were tested in the morning and were rested. In comparison those in the wake condition were tired as they were tested in the evening and as a result they performed worse. One standard solution to rule out this potential confound is to include two additional conditions where the study and test phases (without a 12-hour delay) occur in either the evening or the morning (also see Nemeth et al., 2019). I added time-of-day controls to all experiments using a short retention interval. This enabled me to also confirm that I had the power to detect reasonably-sized effects of well-known differences in memory after immediate and delayed testing. Therefore, for all experimental studies I used four conditions: sleep (PM-AM), wake (AM-PM), AM time-of-day control (AM-AM), and PM time-of-day control (PM-PM).

#### *Online recruitment*

Most often, investigations of sleep on memory are conducted in the lab. This approach arguably allows researchers better control over participants' behaviour at study and test and either the intervening period of sleeping, for the sleep condition, or wakefulness for the wake condition. For example, during study and test phases, researchers can ensure that participants are paying attention to the task to ensure that better quality data is collected. During the intervening period of sleep, researchers can ensure that participants actually sleep between

study and test to maximise the benefits of sleep over a period of wake and obtain objective measures of sleep (e.g., sleep onset, sleep duration, awakenings). During wake however, researchers often allow participants to go about their day and return for testing in the evening.

This approach has drawbacks that can be resolved by conducting experiments online. A prevailing issue facing sleep researchers is small sample sizes. Many sleep and memory experiments, investigating the impact of sleep on memory consolidation collect around 10 to 20 participants per condition. This is problematic because small sample sizes can increase the chance of a Type I error (i.e., falsely rejecting the null hypothesis), which means that the probability of identifying an effect that is true can be lower if the true effect size is much smaller than is reported in the literature because of publication bias (see Button et al., 2013). By collecting data online, power is maximised and sampling errors are reduced thereby reducing the likelihood of Type I errors. This is achieved simply by using the same AM-PM: PM-AM design used by sleep researchers. Instead of bringing participants to the lab they complete the study and test phases using their computers and are presumed to either sleep or remain awake during a 12-hour retention interval. To ensure that participants pay attention during the tasks, attention checks are employed. By asking participants questions about their sleep behaviour during the night (e.g., How long did you sleep last night?) or asking participants about their wake behaviour during the day (e.g., Did you take a nap today?), participants can be included or excluded based on their answer to these questions.

Although some may argue that online experiments yield less reliable data, research indicates otherwise. For example, a number of researchers have found that behavioural effects found in the lab are also replicated when tested online (e.g., Arechar, Gatcher &, Molleman, 2018; Bartneck, Duenser, Moltchanova, & Zawieska, 2015; Kees, Berry, Burton, & Sheehan, 2017). This makes it unlikely that the benefits of sleep would be limited to lab-based studies. Finally, recruiting from online work source sites, such as mTurk ([www.mturk.com](http://www.mturk.com)), yields

datasets that are more generalizable to the population than university undergraduates that are typically recruited for these types of experiments. But it should be noted that the debate about the reliability/generalisability of data collected online is still ongoing (see Hauser, Paolacci & Chandler, 2018). Indeed in our own data we found that the mTurk sample may be more highly educated relative to the general population. And others have indicated that participants recruited online may not put in sufficient effort into the task because they are unsupervised (Ford, 2017). Although such issues can be mitigated (e.g., 95% approval ratings that allow recruitment of the most reliable participants). The critical question is whether such drawbacks are worse than running the risk of carrying out experiments with small samples that may be more likely lead to Type II errors (e.g., Button et al., 2013).

#### *Pre-registration and registered reports*

Many results in psychological science research fail to replicate (Rahal & Open Science Collaboration, 2015). Large scale collaborations that investigated reproducibility in psychological science have found that many results do not replicate, and if they do, the effect sizes were much smaller than the original studies (Rahal & Open Science Collaboration, 2015). One explanation for these un-replicable findings is that they are indicative of publication bias. That is, mostly positive results are published and null results remain in the “file drawer” or on computer backups (Franco, Malhorta & Simonovits, 2014). Researchers have also proposed that original results may not replicate because of questionable research practices (QRPs). QRPs include post-hoc use of analyses and selective reporting of results, hypothesising after results are known, and performing experiments that are underpowered (e.g., John, Loewenstein & Prelec, 2012).

To combat these issues some researchers have encouraged transparency in research, advocating pre-registering hypotheses and analysis pipelines prior to collecting data and

specifying a-priori power analyses to determine sample size (Munafò et al., 2017). Pre-registering can prevent optional stopping of data collection that could lead to false positives and claiming exploratory findings are confirmatory (e.g., selective reporting of analyses that support exploratory outcomes which were not initially predicted) (Forstmeier, Wagenmakers, & Parker, 2017). The former is achieved by encouraging researchers to provide an a-priori power analysis which determines the effect which the researchers can detect with a given sample size (Munafò et al., 2017). The latter is achieved by encouraging researchers to register the confirmatory analyses that will test their pre-registered hypotheses (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). By pre-registering the analyses researchers can be held to account if they diverge from their pre-registrations. Any remaining analyses must be labelled as exploratory. This transparent approach has been adopted by some across the sub-disciplines in psychological science, culminating in a more robust basis for scientific progress where reproducibility and transparency are valued.

However, this approach does not deal with the problem of publication bias. Pre-registering does not make it any less likely that only positive results will be published and that null results are left in the file drawer. One solution is the registered report format which integrates pre-registration into the publication submission process (Chambers, 2013). For example, similar to a pre-registration, prior to beginning data collection participants submit an introduction outlining their theoretical rationale and hypotheses, the methods that will be used and an analysis pipeline which will test the proposed hypotheses. This is known as a Stage 1 submission, which undergoes peer review, and if deemed suitable by the journal editor, the manuscript is accepted “in principle”. In principle acceptance means that if the researchers adhere to the protocol and analyses the manuscript will be accepted for publication at Stage 2 submission. Importantly, this will happen if the results are positive or null. This format also does not stifle exploratory analyses, as researchers are still allowed to conduct these types of



analyses as long as they are labelled as such. This combats publication bias by ensuring that null results are published and that new research questions devised after the fact are also included. In the current thesis, I adopted both pre-registration and the registered report format.

### Summary of the current thesis

#### *Sleep-based memory consolidation and recognition memory*

Based on the evidence that sleep benefits memory in comparison to an equivalent period of wake, I used the methods described previously to conduct investigations into the impact of sleep on recognition memory. The first experiment is presented in Chapter 2 and is an investigation of the impact of sleep on eyewitness identifications. It is published as a registered report (Morgan, Tamminen, Seale-Carlisle & Mickes, 2019). At the time of in principle acceptance it was the first experiment to investigate the role of sleep in eyewitness identifications. After witnessing a crime, the witness may be presented with a lineup to try to identify the perpetrator. A lineup is an array of photographs including the police suspect (who is innocent or guilty) and fillers (who are known to be innocent). Given that sleep in comparison to a period of wake yields greater recognition memory performance (e.g., Drosopoulos, Wagner & Born, 2005; Jones, MacKay, Mantua, Schultz & Spencer, 2018; Wilson, Baran, Pace-Schott, Ivry & Spencer, 2012), I predicted that sleep would improve eyewitness ID performance. Specifically, I predicted that sleep would improve discriminability (i.e., the ability to discriminate between innocent vs. guilty suspects) anticipating that sleep would play an active role in consolidating memory for the perpetrator, although it should be acknowledged that in this experiment we were unable to distinguish between an active or passive role of sleep. Another type of accuracy, reliability (i.e., the likelihood that the identified suspect is guilty), was also measured, but I had no strong predictions regarding the impact of sleep. In the largest experimental sleep-recognition memory experiment to date ( $N = 4000$ ), participants were

shown a mock crime, slept overnight or remained awake during the day and were tested on their memory for the perpetrator.

Whilst data collection was underway, I conducted another experiment examining the impact of sleep-based memory consolidation on the misinformation effect (Chapter 4). The misinformation effect occurs when an individual encounters false information about an event or situation that they later endorse as having experienced (Loftus, 1975). It is easy to appreciate how this might be problematic (e.g., reduction in vaccination rates has led to increases in measles; Majumder et al., 2015; Public Health England, 2018), therefore finding ways to mitigate this effect is worthwhile.

Sleep could be considered as one source of mitigation, but the evidence is mixed. One study indicated that when participants are presented with an event and encounter misinformation for that event after a period of sleep their memory for the event was better compared to wake, but there was no difference in the endorsement of misinformation (van Rijn, Carter, McMurtrie, Wilner, & Blagrove, 2017). Another experiment, using a similar design, showed that participants in the sleep condition had greater memory for the original event compared to the wake condition, but they also endorsed more misinformation compared to the wake group (Calvillo et al., 2016).

In both experiments, sleep benefitted memory for the original information. However, the results differed regarding whether sleep leads to a greater misinformation effect. Calvillo et al. (2016) concluded that the misinformation effect was greater in the sleep condition because sleep stored gist-based representation of the original event. However, this conclusion does not fully follow the pattern of results because memory was greater for the sleep condition compared to the wake condition. Additionally, based on the existing sleep literature, one might expect

that sleep would aid consolidation of the original event and should enable participants to better encode the misinformation and in turn reject it at test.

Calvillo et al. (2016) also suggested that participants remembered that they were presented with original information and misinformation, but were not able to remember the source. Based on the existing sleep and memory literature it seemed more likely that sleep would lead to a greater misinformation effect if the misinformation is encountered before sleep rather than after (e.g., Cairney et al., 2018). Broadly speaking, I predicted that sleep would benefit recognition memory, lead to greater memory for misinformation and would reduce source memory relative to a period of wake. Clearly one way of resolving the aforementioned discrepancies was to conduct a conceptual replication with a source memory task, and position misinformation before and after a period of sleep or wake, which I carried out and presented in Chapter 4. Unlike the experiment presented in Chapter 2 an active role of sleep in memory consolidation could be examined. For example if sleep has an active involvement in memory consolidation it might be expected that sleep consolidation would integrate true information and misinformation – which is essential for updating pre-existing representations of information.

The registered report presented in Chapter 2 and the experiment presented in Chapter 4 did not yield convincing evidence to indicate that sleep impacts memory consolidation. In the registered report (Chapter 2) contrary to our predictions sleep did not benefit discriminability despite the fact that we were powered to detect a very small effect ( $d = .180$ ). We explored numerous possible explanations for the discrepancy between our findings including differences in recognition memory tasks, sleepiness between study and test, etc. Whilst data collection was underway another experiment indicated that sleep reduces false identifications (i.e., incorrectly identifying an innocent suspect; in this case a filler who is known to be innocent) in target-absent lineups (Stepan, Dehnke, & Fenn, 2017). However, we were also unable to replicate

their findings in our dataset and argue that their finding is likely to be a false positive. Sleep also did not impact reliability.

The findings in the experiment presented in Chapter 4, examining the impact of sleep on the misinformation effect, also yielded a peculiar set of exploratory results that should be interpreted with caution. Namely in this experiment there was no evidence of forgetting between experimental and control conditions. That is, there was no difference in old/new recognition memory (i.e., discriminating targets from lures), memory for misinformation (i.e., discriminating misinformation from lures), and source memory (i.e., discriminating true information from misinformation), between conditions with a 5-minute retention interval and a 12-hour retention interval. These results indicated that there may be something problematic with the paradigm used in my experiment. Therefore, I cautiously interpreted our results. The result that misinformation presented after sleep leads to a greater misinformation effect also did not conceptually replicate. This did not vary depending on whether the misinformation was presented before or after the retention interval. Only one confirmatory analysis yielded a significant difference between sleep and wake conditions; one sleep condition had greater old/new recognition memory performance than the wake condition.

Taken together these findings paint a different picture relative to the evidence that is published. For instance, in both experiments we were powered to detect the effect that sleep would benefit memory consolidation in a recognition memory task. To my surprise I did not find consistent evidence that sleep aids memory consolidation despite using adequate sample sizes. In response to this I re-examined the literature investigating the impact of sleep on recognition memory over a twelve-hour period. Interestingly, although sleep does appear to benefit recognition memory, the way in which it appears to do so is not consistent. The question is whether inconsistencies reflect selectivity of sleep based memory consolidation as some have

argued (e.g., Diekelmann, Wilhelm & Born, 2009; Feld, & Diekelmann, 2015) or whether these inconsistencies are indicative of a possible publication bias.

*Sleep and emotional memory: A conceptual replication attempt*

Forming a basis for confirming if the benefits of sleep on memory are selective or whether the published findings reflect publication bias requires replication of previous research. Replication enables researchers to confirm or disconfirm our current understanding of the impact of sleep on recognition memory. By systematically attempting to replicate previous studies researchers can eliminate findings which may have been false positives and simultaneously advance our theoretical understanding of sleep and memory, such as if the sleep based memory consolidation process is selective. One strategy for introducing a replication effort into a field where replication is not common is by targeting an area in which inconsistency is already apparent. One line of investigation in which there is already inconsistency within the sleep and memory literature is the impact of sleep on emotional memory, which I re-examined in Chapter 5.

Sleep is thought to selectively consolidate emotional memories over neutral memories (Hu, Stylos-Allan & Walker, 2006; Nishida et al., 2009; Payne et al., 2008; Wagner et al., 2001). Researchers have argued that at encoding emotional memories are “tagged” so that they are selectively preserved during the memory consolidation processes that occur during sleep (van der Helm & Walker, 2011). However, the evidence supporting this argument is mixed. Some find that sleep compared to a period of wake preserves memory for negative images (e.g., Hu, Stylos-Allan & Walker, 2006; Nishida et al., 2009; Payne et al., 2008; Wagner et al., 2001), while others find no effect of emotion (e.g., Baran, Pace-Schott, Ericson & Spencer, 2012; Cellini, Torre, Stegagno & Sarlo, 2016; Lehmann, Schreiner, Siefrietz & Rasch, 2016; Lewis, Cairney, Manning & Critchley, 2011). The discrepancies between these findings do not appear

to be due to varying designs, as many of them are similar. Therefore, a conceptual replication was required.

I also tested an alternative theoretical explanation for the inconsistencies in the findings in this research. That is the inconsistencies may be indicative of the varying impact that sleep has at different levels memory strength. To this end, I present evidence that points towards this, even outside of the literature examining the impact of sleep on recognition memory. For example, sleep has been shown to benefit free recall, more so than associative memory and recognition memory (e.g., Schönauer et al., 2013). Sleep has also been found to benefit highly rewarded items and not items associated with low reward (e.g., Feld, Besedovsky, Kaida, Münte, & Born, 2014). For example, some studies indicate that sleep benefits familiarity-based memory and not recollection (e.g., Daurrat et al., 2007; Drospolous et al., 2005), whereas others indicate the opposite (e.g., Atienza & Cantero, 2008). Another finding is that sleep benefits gist-based memory (e.g., Payne et al., 2009) and not verbatim-based memory. These findings are conflicting because it appears that sometimes sleep benefits strong memories (e.g., free recall, high reward, recollection) and other times it benefits weak memories (e.g., recognition, familiarity, gist) at retrieval. One can also easily appreciate that emotional memories might be strong and neutral memories might be weaker memories. One way of measuring the impact of sleep on memory strength is by measuring changes in the relationship between confidence (i.e., a proxy for memory strength) and accuracy. For instance, if sleep benefits strong memories over weaker memories, then accuracy should be greater for high confidence memories in the sleep condition compared to the wake condition.

In Chapter 5 I present an experiment investigating the impact of sleep on emotional memory and assessing the replicability of this effect. I used the same AM/PM PM/AM design as the previous experiments, except that participants were presented with negative and neutral images at study and were tested on their memory for those images. I expected to find that sleep

would benefit recognition memory compared to a period of wake, and that participants would remember negative images more so than neutral images. I had no strong predictions about the interaction between sleep and emotional memory given the mixed literature, and had no strong predictions about the impact of sleep on memory strength. If sleep had an active benefit on memory consolidation it would be expected that sleep would selective enhance emotional memories over neutral memories to a greater magnitude compared to an equivalent period of wake. Sleep did not selectively benefit negative images over neutral images and negative images were not better remembered compared to neutral images. However, sleep did yield a significant recognition memory advantage compared to wake and benefited high memory strength recognition judgements (for new, but not old, items), compared to wake. However, these findings are difficult to interpret considering potential time-of-day effects.

*Norms for sleep researchers and clinicians: Making the most of a “bad” situation*

It is easy to appreciate that the experiments discussed thus far do not present an ideal pathway for career progression as an academic under the current publish or perish culture (Miller, Taylor, & Bedeian, 2011). As previously acknowledged, although these experiments were pre-registered only one was a registered report, and only that one experiment is guaranteed to be published. Fortunately, as part of that registered report we collected an abundance of data from over 7,000 individuals about their sleep related behaviour. Very little normative data exist for sleep related behaviour and therefore, in Chapter 3 we compile and analyse these data to provide normative data for the Stanford Sleepiness Scale, Epworth Sleepiness Scale and sleep duration. It should be pointed out that the sample analysed in the registered report compared to Chapter 3 is much smaller. This is because hundreds of participants were excluded from the analyses in the registered report due to inclusion criteria and thousands of participants only completed the first part of the experiment. Despite this data were still available from those participants that could be used to provide norms for the

aforementioned scales. Using these data, we examined the impact of gender, age, education and time of day on all scales, finding that only the Stanford Sleepiness Scale is impacted by time of day. Participants report being sleepier in the evening than the morning. This may be an unidentified confound in many experiments using the AM PM design as many of them are not powered to detect this effect. This may also add to scepticism surrounding the research examining the impact of sleep on recognition memory.

### *Aims of the current thesis*

The aims of this thesis were incremental. I first sought to conduct a large-scale examination of the impact of sleep on eyewitness identifications alongside a re-examination and extension of the research investigating the impact of sleep on the misinformation effect. Given that neither of these experiments yielded strong evidence to indicate that sleep benefits recognition memory, we pursued a replication effort re-examining the impact of sleep on emotional memory. Lastly, although the data examining the impact of sleep on recognition memory are unclear, exploratory analyses of the questionnaire data collected in our registered report allowed us to provide norms for clinicians and researchers. Overall, my aim is that the findings and methods presented within this thesis will motivate a movement for open science and reproducibility in the sleep and memory research community.



## Chapter 2: Examining the Impact of Sleep on Eyewitness Identifications

### Abstract

Sleep aids the consolidation of recently acquired memories. Evidence strongly indicates that sleep yields substantial improvements on recognition memory tasks relative to an equivalent period of wake. Despite the known benefits that sleep has on memory, researchers have not yet investigated the impact of sleep on eyewitness identifications. Eyewitnesses to crimes are often presented with a lineup (which is a type of recognition memory test) that contains the suspect (who is innocent or guilty) and fillers (who are known to be innocent). Sleep may enhance the ability to identify the guilty suspect and not identify the innocent suspect (i.e., discriminability). Sleep may also impact reliability (i.e. the likelihood that the identified suspect is guilty). In the current study, we manipulated the presence or absence of sleep in a forensically relevant memory task. Participants witnessed a video of a mock crime, made an identification or rejected the lineup, and rated their confidence. Critically, some participants slept between witnessing the crime and making a lineup decision, while others remained awake. The prediction that participants in the sleep condition would have greater discriminability compared to participants in the wake condition was not supported. There were also no differences in reliability.

## Introduction

Based on a large body of neuroscientific and behavioural evidence, conventional wisdom holds that sleep helps people remember better. While sleeping, newly learned information strengthens and stabilises (i.e. consolidates) as a result of continued processing occurring in neural circuits which are well-established to be critical for memory (Diekelmann & Born, 2010; Rasch & Born, 2013; Stickgold, 2005). This is demonstrated time and time again in behavioural experiments in which the results indicate that participants who sleep between learning information and being tested on that information consistently outperform participants who do not (Backhaus, Hoeckesfeld, Born, Hohagen & Junghanns, 2008; Gais, Molle, Helms & Born, 2002; Gui et al., 2017; Horvath, Myers, Foster & Plunkett, 2015; Payne et al., 2012; Plihal & Born, 1997; Rasch, Buchel, Gias & Born, 2007; Tamminen, Payne, Steikgold & Wamsley, 2010). This finding applies widely on tests of recognition memory for words, objects, faces, and locations (Drosopoulos, Wagner & Born, 2005; Maurer, Zitting, Elliott, Czeisler, Ronda, Duffy, 2015; Payne, Stickgold, Swanberg, Kensinger, 2008; Sheth, Nguyen, Janvelyan, 2009; Wager, Kashyap, Diekelmann & Born, 2007; Wilson, Baran, Pace-Schott, & Ivry, 2012). Theoretically this should also apply to memory for crimes. However, the role that sleep may play on eyewitness identification has yet to be empirically investigated.

During the course of a criminal investigation, once the police find a suspect, a lineup may be administered to an eyewitness. A lineup is a type of recognition memory test that helps police assess the likelihood that the suspect is indeed the offender. Lineup procedures generally involve presenting the eyewitness with the police suspect amongst fillers. Fillers are known to be innocent, and physically resemble the suspect or match the description of the offender (Police and Criminal Evidence Act 1984, Code D, 2011; Technical Working Condition for Eyewitness Evidence, 2003). Of the three possible responses an eyewitness can make – identify

the suspect, identify a filler, or identify no one, whether or not the suspect was identified is typically of most applied interest. In this regard, if the suspect in the lineup is guilty, and the eyewitness chooses him or her, then that is a correct identification, and if the lineup is rejected, then that is a miss. If, on the other hand, the suspect in the lineup is innocent, and the eyewitness chooses him or her, then that is a false identification, and if the lineup is rejected, then that is a correct rejection. To put another way, two outcomes are errors (misses and false identifications) and two outcomes are accurate (correct identifications and correct rejections).

Considerable research has been conducted on a variety of variables that affect eyewitness memory performance (Carlson & Carlson, 2014; Carlson, Dias, Weatherford & Carlson, 2016; Clark, Brower, Rosenthal, Hicks & Moreland, 2013; Clark, Marshall & Rosenthal, 2009; Colloff, Wade & Strange, 2016; Dobolyi & Dodson, 2013; Dodson & Dobolyi, 2016; Dodson & Dobolyi, 2016; Greathouse & Kovera, 2009; Gronlund et al., 2012; Gronlund, Carlson, Dailey & Goodsell, 2009; Haw & Fisher, 2004; Hulse & Memon, 2006; Juslin, Olsson & Winman, 1996; Lindsay & Wells, 1985; Luus & Wells, 1991; Malpass & Devine, 1981; Malpass & Kravitz, 1969; Mickes, 2015; Mickes, Flowe & Wixted, 2012; Molinaro, Arndorfer & Charman, 2013; Palmer & Brewer, 2012; Palmer, Brewer, Weber & Nagesh, 2013; Palmer, Brewer, Weber & Nagesh, 2013; Pickel, 1998; Porter, Moss & Reisberg, 2014; Read, Yuille & Tollestrup, 1992; Sauer, Brewer, Zweck & Weber, 2010; Sauerland et al., 2016; Schooler & Engstler-Schooler, 1990; Smith & Flowe, 2014; Tunnicliff & Clark, 2000; Valentine & Mesout, 2008; Wells, Rydell & Seelau, 1993; Wetmore et al., 2015; Wilson, Seale-Carlisle & Mickes, 2018; Wright, Boyd & Tredoux, 2001). How these variables affect eyewitness identifications are of interest to two different types of decision makers: policymakers (such as Police and Crime Commissioners and Police Chiefs) and triers of truth (such as judges, magistrates, and jurors). And because they have different decisions to make, they are consumers of different types of analyses, both of which entail suspect

identifications (Mickes, 2015). One type of analysis assesses discriminability, which is the ability for eyewitnesses to distinguish innocent from guilty suspects. Procedures that reduce false identifications and increase correct identifications are procedures that have better discriminability than those that do not. Instituting procedures that increase discriminability is in the remit of policymakers. The other type of analysis assesses reliability, which is the probability that the identified suspect is the offender. In assessing the likelihood that a defendant is guilty, reliability, not discriminability, is informative to triers of truth. While the different assessments of discriminability and reliability both use suspect identifications, it is possible for the outcomes to diverge.

A variable low in discriminability may give rise to high reliability. Take, for example, a study in which exposure duration was manipulated; participants saw the target for either 5 seconds or 90 seconds (Palmer et al., 2013). Discriminability was unsurprisingly lower for participants in the 5-second condition compared to participants in the 90-second condition, but identifications made with high confidence were equally high in accuracy regardless of condition<sup>2</sup>. In other words, these identifications were reliable whether the target was seen for 5 or 90 seconds. Other variables that have also shown this pattern of results where discriminability is lower for one condition than another, but reliability is equivalent (for identifications made with high confidence) include: sequential versus simultaneous lineups (Mickes, 2015), short versus long retention intervals (Wetmore et al., 2015), weapon present versus weapon absent (Carlson et al., 2014; Carlson et al., 2016), offenders and witnesses are of different versus same races (Dodson & Dobolyi, 2016), and a verbal description versus no verbal description was provided (Wilson, Seale-Carlisle & Mickes, 2018).

---

<sup>2</sup> Focus is typically placed on identifications made with high confidence because these are the identifications that matter in the court of law.

Unlike the cases above in which discriminability and reliability diverge, a variable low in discriminability may also be low in reliability. Take, for example, a study in which US lineups were compared to UK lineups. US lineups yielded higher discriminability and higher reliability than UK lineups (Seale-Carlisle & Mickes, 2016). This pattern also emerged when showups (i.e. just the police suspect, no fillers, are presented) were compared to simultaneous lineups. In these studies, discriminability and reliability were both higher when memory was tested on a lineup compared to when it is tested on a showup (Wetmore et al., 2015; Mickes, 2015). Similarly, unfair lineups, compared to fair lineups, were lower in discriminability and reliability (Collof et al., 2015, Grondlund et al., 2012, Mickes, 2015). Thus, a variable low in discriminability can be high or low in reliability, and (though we know of no research showing this is the case) it is theoretically possible that a variable high in discriminability can be low in reliability (Mickes, 2015). If sleep impacts discriminability, then scheduling identification procedures at optimal times would be in the purview of policymakers. If sleep impacts reliability, then knowing if sleep, and the quality of sleep that occurred between the crime and the ID would be of interest to triers of truth.

How sleep affects witnesses' ability to identify an offender from a lineup is currently unknown. While some attention has been paid to the effects of sleep deprivation on eyewitness memory (Blagrove & Akehurst, 2000; Frenda, Pathis, Loftus, Lewis & Fenn, 2014), as far as we know, only one study correlated aspects of sleep with recognition memory for a mock crime (Thorley, 2013). In this study, participants 1) watched a mock crime video, 2) completed questionnaires about sleepiness, sleep quality, and sleep duration, and 3) were immediately tested on a recognition memory test of central and peripheral details of the video. Sleepiness and poor sleep quality were negatively correlated with the accuracy of recall of the peripheral details, but not the central details. This result provides some evidence that sleep affects some aspects of eyewitness memory, however, participants were not asked to identify the perpetrator.

Furthermore, confidence ratings were not collected, and hits and false alarms were not reported, so the impact sleep has on discriminability and reliability remains unknown. Given the well-known beneficial effects that sleep has on memory and the lack of knowledge about the impact that sleep has on eyewitness identifications, investigations of this type are highly overdue.

The goal of the current study was to measure the impact of sleep on discriminability and reliability on experimental eyewitnesses. To do so, we combined the AM-PM PM-AM sleep design with a forensically relevant design. The AM-PM PM-AM design is commonly used to compare effects of sleep versus wake on recognition memory (Maurer et al., 2015; Payne et al., 2008; Sheth et al., 2009; Tamminen et al., 2010). In this design, participants are assigned to a wake (AM-PM) or sleep (PM-AM) condition. In a forensically relevant design, which is similar to a standard list-learning recognition memory test, participants take part in a study (encoding) phase and test (retrieval) phase. A video of a mock crime is viewed in the former and memory for the target in the video is tested on a lineup (either target-present or target-absent) in the latter. Unlike in a standard list-learning recognition memory test where there are multiple trials per participant, there is only one trial per participant in a forensically relevant design (to mimic the experience of a real eyewitness). In our integrated design, participants in the wake condition took part in the study phase in the morning (AM) and in the test phase in the evening (PM). Participants in the sleep condition took part in the study phase in the evening (PM) and in the test phase in the morning (AM). To rule out any influence of time-of-day effects on discriminability and reliability which may otherwise explain our findings, we included two circadian control conditions where both the study phase and test phase occurred either in the AM or in the PM to assess potential time-of-day confounds.

Based on the existing sleep literature, we predicted that discriminability will be greater for those in the sleep condition compared to those in the wake condition. Previous literature on sleep and declarative memory show that sleep benefits are not explained by the test phase

occurring at a different time of day in the sleep and wake conditions (e.g., Sheth et al., 2009; Barrett & Ekstrand, 1972; Ellenbogen, Hulbert, Stickgold & Dinges, 2006) and so we predicted no difference between the AM-control and PM-control conditions for discriminability. Due to the lack of literature of the effects of sleep and time-of-day effects on reliability, we did not have sufficient grounds to make predictions about this measure. However, because few variables appear to have an appreciable effect on high-confidence accuracy (Wixted & Wells, 2017), it seems reasonable to suppose that the same might be true of sleep. This investigation is a first step toward understanding how sleep may influence eyewitness identification performance.

## Methods

### *Participants*

We used a sample size similar to other forensically relevant experiments in which discriminability and reliability were assessed (i.e.  $n = 1000$  per condition) (Colloff et al., 2016; Sealse-Carlisle & Mickes, 2016). Sensitivity analysis revealed that by setting the parameters to standard values of  $\alpha = 0.0125$  (corrected for each hypothesis test  $\alpha = 0.05/4 = 0.0125$ ) and  $1 - \beta = 0.95$ ,  $n = 1000$  per experimental condition can detect an effect size of  $d = 0.18$  in a two-tailed test. Participants ( $N = 4000$ ) were randomly assigned to one of four conditions: wake ( $n = 1000$ ), sleep ( $n = 1000$ ), AM control ( $n = 1000$ ), and PM control ( $n = 1000$ )<sup>3</sup>. Participants were also randomly assigned to a target-present or target-absent lineup. When 4000 participants took part, data collection ceased. The Royal Holloway, University of London Research Ethics Committee approved this study.

---

<sup>3</sup>Participants were randomly assigned to either the AM control or wake condition and the PM control and sleep condition depending on the time of day they chose to participate at.

Participants were recruited online using the work source site Amazon Mechanical Turk ([www.mturk.com](http://www.mturk.com)). To ensure good quality data, participation was limited to those who have hit approval rates of at least 85%. Approval rates are based on the quality of the responses on past tasks. Participant location was restricted to the US to ensure that data collection took place across only four time zones (Eastern Standard Time, Central Standard Time, Mountain Standard Time, and Pacific Standard Time). Individuals had to meet the following inclusionary criteria: be between the ages of 18-40 years, not currently diagnosed with any sleep disorder (e.g. insomnia, sleep apnoea, etc.), psychiatric disorder (e.g. depression, posttraumatic stress disorder, etc.), or neurological disorder (e.g. mild cognitive impairment, Alzheimer's disease, etc.) which affects memory (Ciipolli, Mazzetti & Plazzi, 2013; Golier et al., 2002; Hornung et al., 2008; Koen & Yonelinas, 2014; Pace-Schott, Germain & Milad, 2015), not work as a shift worker, have not travelled across time zones within two weeks prior to participating, and not currently taking any prescribed medication that may affect sleep or memory.

### *Materials*

*Pre-screening questions.* The pre-screening questions included yes-no questions about the inclusion criteria listed above.

*Video.* The study material was a 35-second video featuring a young adult White male (the target) stealing a laptop and mobile phone from an empty office. The target's face was in clear view throughout the video.

*Lineups.* The lineups were 6-person simultaneous photo lineups. Target-present lineups contained a photo of the target and 5 fillers, and target-absent lineups contained 6 fillers. The fillers matched the description of the target, which were provided by participants ( $n = 19$ ) who watched the video and answered questions about the target's appearance. The averaged



descriptors were entered into the Florida Department of Corrections database ([www.dc.state.fl.us](http://www.dc.state.fl.us)), and photos of 100 individuals who matched were extracted and grey-scaled. The target and fillers were randomised to appear in any of the six lineup positions.

*Distractor Task.* The distractor task was anagram puzzles (Colloff et al., 2016; Mickes et al., 2012). Participants tried to solve 50 anagrams of US states for 5 minutes. This task was relevant to participants in the control conditions, however all participants completed this task for the sake of consistency. Distractor tasks are standard in forensically relevant designs to prevent rehearsal between witnessing the crime and making an identification (Colloff et al., 2016; Mickes et al., 2012).

*Sleep related questions.* Participants in the wake condition were instructed not to nap between the study and test phases, and were asked at the beginning of the test phase whether they did nap. If they answered yes, their data were excluded from the analysis and were replaced, as napping has been shown to aid memory consolidation (Lahl, Wispel, Willigens & Pietrowsky, 2008). Participants in all conditions were asked when they went to sleep, woke up, and to estimate how long they slept prior to participating in the study phase. Participants in the sleep condition were asked those questions again for their night's sleep between the study and test phase.

*Stanford Sleepiness Scale (SSS).* The SSS is used to measure an individual's current sleepiness using a 7-point scale (1 = feeling active, vital, alert or wide awake; 7 = no longer fighting sleep, sleep onset soon; having dream like thoughts) (Hoddes, Zarcone, Smythe, Phillips & Dement, 1973). High scores indicate high levels of sleepiness and low scores indicate low levels of sleepiness. The data collected from the SSS was used as a manipulation check (see analysis strategy) to ensure the sleep and wake conditions differed only on this variable and not on tiredness at the time of participation.

*Reduced Morningness Eveningness questionnaire (MEQr).* The MEQr measures the time of day an individual is more alert (chronotype) by answering five questions about their sleep habits (e.g. “At what time of the day do you feel you become tired as a result of need for sleep?”) (Adan & Almirall, 1991). Low scores indicate evening types and high scores indicate morning types. The data collected from this instrument were used for exploratory analyses only.

*Epworth Sleepiness Scale (ESS).* The ESS measures the general level of daytime sleepiness. Participants rate their chance of dozing while completing 8 everyday tasks on 4-point scale (0 = would never doze, 1 = slight chance of dozing, 2 = moderate chance of dozing, 3 = high chance of dozing) (Johns, 1991). ESS scores range from 0-24. Low scores indicate normal levels of daytime sleepiness and high scores indicate high levels of daytime sleepiness. Participants who scored between 16-24 on the ESS (possibly indicating sleep difficulties) were excluded from the analysis and replaced.

*St Mary’s Hospital Sleep (SMHS) questionnaire.* The SMHS questionnaire measures duration and quality of the prior night’s sleep (Ellis et al., 1981). Participants answered, “How well did you sleep last night?” by responding on a 6-point scale (1 = very badly; 6 = very well). Low scores indicate poor sleep quality and high scores indicate high quality of sleep. The data collected using the SMHS questionnaire were used for exploratory analyses only.

*Procedure.* Participation took place online. Participants first answered the pre-screening questions. Those who did not meet the inclusionary criteria did not proceed to the experiment. In the study phase, participants digitally signed the consent form, answered the sleep-related questions, completed the SSS, watched the mock crime video, and completed the distractor task. In the test phase, participants tried to identify the target from the video in a lineup. Confidence ratings on a scale from 0-100% (0 = guessing, 100% = certain) were made

for lineup decisions. Then, participants were asked if they have consumed any caffeine between the study and test phase. This information was analysed as part of our exploratory analyses. Next, participants completed the SSS, MEQr, ESS, and the SMHS, in fixed order. Participants in the sleep condition answered the sleep-related questions a second time but participants in the wake and control conditions did not (as they did not sleep between study and test). After completing the sleep questionnaires and sleep-related questions participants answered a validation question about the video (“What was the crime committed in the video?”) to assure that attention was paid during the study phase. If answered incorrectly, the data were excluded from analysis<sup>4</sup>. The test phase, sleep questionnaires and validation question were self-paced. Finally, participants were debriefed.

The order of the study and test phase was consistent across participants, but the timing varied. The procedure occurred at two different times for the participants in the sleep and wake conditions and only once for the control participants. Participants in the wake condition completed the study phase in the morning (between 8 a.m. and 11 a.m.<sup>5</sup>) and completed the test phase approximately 12 hours later in the evening (between 8 p.m. and 11 p.m.). Participants in the sleep condition completed the study phase in the evening (between 8 p.m. and 11 p.m.) and completed the test phase approximately 12 hours later in the morning (between 8 a.m. and 11 a.m.). Participants in the AM control condition completed study and test phases in the morning (between 8 a.m. and 11 a.m.), and participants in the PM control condition completed study and test phases in the evening (between 8 p.m. and 11 a.m.). The experiment was available for participation only during these hours.

---

<sup>4</sup> Previous research testing online with this type of design shows that less than 2% of participants answer this incorrectly (e.g., [57]).

<sup>5</sup> As approved by the editor during data collection after the Stage 1 manuscript was accepted, the timings for data collection were extended from 8 a.m. - 10 a.m. and 8 p.m. - 10 p.m. to aid recruitment.

## Analysis Strategy

Participants who incorrectly answered the validation questions, participants who reported napping between study and test phases, participants who scored between 16 to 24 on the ESS (possibly indicating sleep difficulties) and participants who reported having less than 6 hours of sleep (Sheth et al., 2009) before the test phase were excluded from all analyses. Participants who were excluded were replaced to achieve the desired sample size ( $N = 4000$ ). Alpha levels were set to 0.05 and Bonferroni corrections were used for multiple comparisons.

### *Correct and false ID rates*

Correct ID rates were computed by dividing the number of guilty suspects identified by the total number of target-present lineups. Because there is no designated innocent suspect, the false ID rate was estimated, which is standard practice (Palmer et al., 2013). The estimated false IDs were computed by dividing the number of innocent suspects identified by the number of lineup members (6). Then this value is transformed into the estimated false ID rate by dividing it by the total number of target-absent lineups.

### *Discriminability*

Receiver operating characteristic (ROC) analysis was conducted to measure discriminability (National Research Council, 2014; Wixted & Mickes, 2012). The most common ROC approach is to collect eyewitness' confidence in their identifications and plot correct ID rate and false ID rate pairs for every level of confidence, and measure the area under the ROC curve (Gronlund, Wixted & Mickes, 2014). Because the false ID rate range extends from 0 to a value less than 1, we conducted partial area under the curve (pAUC) analysis. To test our hypothesis that discriminability will be higher for those in the sleep condition compared

to those in the wake condition, we compared the pAUC values from these conditions. These analyses were conducted using the pROC package in R (Robin et al., 2011).

To measure pAUCs, a false ID cut-off needs to be specified, and thus we used the most conservative overall false ID rate (i.e., we used the rightmost point on the ROC from the condition that yielded the more conservative responding overall) so the package did not have to extrapolate these points if a larger cut-off was used) (Wixted & Mickes, 2012). The pROC package uses the number of suspect identifications for target-present lineups and target-absent lineups for every level of confidence and plots the empirical ROC as computed using trapezoids and were standardised with the formula

$$\frac{1}{2} \left( \frac{(1 + pAUC - min)}{max - min} \right)$$

where *min* is the pAUC of chance responding and *max* is the pAUC of perfect performance. This was computed for each condition.

To compare the pAUC values, pROC derives *D*. *D* is defined by

$$D = \frac{(pAUC_{sleep} - pAUC_{wake})}{s}$$

where *s* is the standard error of the differences between the two pAUCs estimated by the bootstrap method (the number of bootstraps set to 10,000). The procedure or variable with the greatest pAUC had the better discriminability.

### *Reliability*

Confidence-accuracy characteristic (CAC) analysis was conducted to measure reliability (Mickes, 2015). For CAC analysis, for each level of confidence, the proportion correct was computed (e.g., high confidence guilty suspect IDs divided by the sum of high confidence correct IDs and high confidence estimated false IDs). A bootstrap procedure was

used to estimate the standard errors associated with suspect ID accuracy for each level of confidence for each condition. Observed data on target-present and target-absent lineups were randomly sampled with replacement to obtain a bootstrap sample for each trial. This was repeated for 10,000 bootstrap trials and the standard deviation of these trials yielded the estimated standard error. This was performed separately for each condition. Non-overlapping error bars signify difference between conditions (Seale-Carlisle & Mickes, 2016).

### *Time-of-day Effects*

The control conditions were included to control for time-of-day effects. To determine whether time-of-day effects impact discriminability and reliability for eyewitness identifications the ROCs and CACs for the AM control condition and PM control condition were compared. To rule out differences in discriminability based on differences in circadian rhythms, we compared pAUC values of the two control conditions. Based on previous work, we did not expect these to differ (Barrett & Ekstrand, 1972; Ellenbogen et al., 2006; Sheth et al., 2009). ROC and CAC analysis were conducted using exactly the same procedures as outlined above. If no ROC and CAC differences existed between the AM and PM control conditions, then any differences in the experimental conditions would not be explained by time of day effects.

### *Current Sleepiness*

SSS values from the sleep and wake conditions at study phase and test phase were compared against each other as a manipulation check, using independent-samples t-tests ( $\alpha$  - level set at 0.05, two-tailed), to rule out the impact of sleepiness at encoding and retrieval. Based on existing literature using the same AM-PM design we did not expect statistically significant differences between the conditions at either phase (Maurer et al., 2015; Payne et al.,

2012; Payne et al., 2008; Sheth et al., 2009). Any remaining analyses were labelled as exploratory.

## Results

The archived data are available in the supplementary files and the approved Stage 1 protocol is available at <https://osf.io/x9j87>. In total data from 4309 participants were collected. Of those participants 309 were excluded from all analyses for answering the validation question incorrectly ( $n = 115$ ), napping between study and test ( $n = 36$ ), having less than 6 hours sleep ( $n = 145$ ,  $M = 7.59$  hours,  $sd = 0.96$ , range = 6-12 hours) or scoring greater than 15 on the ESS

**Table 1.** Demographic information by sleep, wake, AM control and PM control groups.

	Group			
	Sleep	Wake	AM Control	PM Control
Gender				
Female	586	604	612	612
Male	410	391	386	384
Do not wish to state	4	5	0	4
Age				
Years	29.55 (5.71)	30.40 (5.36)	30.41 (5.31)	29.43 (5.69)
Ethnicity				
African American	70	67	85	94
Arab American	6	4	3	6
Asian American	85	63	59	67
Caucasian American	720	764	752	704
Hispanic American	65	48	56	61
Indian American	13	17	15	18
Native American	4	2	2	5
Do not wish to state	11	11	10	15
Other	26	24	18	30
Education				
Bachelor's Degree	425	450	415	461
High School/GED	67	83	87	66
Master's Degree	162	154	137	129
Post-Masters	60	60	50	53
Some College	282	249	303	287
Some High School	1	1	3	1
Do not wish to state	3	3	5	3

( $n = 27$ )<sup>6</sup>. The data from the remaining 4000 participants were included in the analyses. Table 1 shows the demographic information<sup>7</sup>.

### *Correct and false ID rates*

Table 2 shows the frequencies of all of the response types by level of confidence for target-present and target-absent lineups for the sleep and wake conditions. The correct ID rates for the sleep and wake conditions were both .79. The estimated false ID rates for the sleep and wake conditions were .07 and .08, respectively.

**Table 2.** Frequencies of correct IDs (CID), filler IDs (FID), and no-IDs for target-present and target-absent lineups for each level of confidence for sleep and wake groups.

Confidence	Sleep					Wake				
	Target-present			Target-absent		Target-present			Target-absent	
	CID	FID	no ID	FID	no ID	CID	FID	no ID	FID	no ID
0	2	0	3	1	4	1	2	3	2	11
10	0	4	1	5	2	1	1	0	6	2
20	2	3	3	13	4	4	1	3	11	9
30	6	7	7	18	15	6	10	6	20	10
40	8	5	9	21	15	16	4	3	27	8
50	11	2	10	30	27	15	5	9	31	16
60	37	11	9	40	37	32	11	12	39	37
70	65	6	14	45	70	46	3	10	51	52
80	69	2	3	25	51	76	5	8	23	51
90	111	2	2	16	28	94	0	5	12	36
100	90	1	2	3	23	120	1	4	3	26
Total		507		493		517			483	

<sup>6</sup> This number does not sum to 4309 because some participants could have been excluded for multiple reasons.

<sup>7</sup> A reviewer of the Stage 2 manuscript recommended that we conduct analysis on differences in demographics between conditions. There were no significant differences in ethnicity [ $\chi^2(12) = 1.55, p = .100$ ] (chi-square limited to ethnicities with cells  $> 5$ ), gender [ $\chi^2(3) = .61, p = .611$ ], or education [ $\chi^2(12) = 1.61, p = .080$ ]. There were significant differences in age between conditions [ $F(3,3996) = 9.18, p < .001$ ], but that is because the sample size is so large. The effect size is trivially small (.007). An age difference of .85 (30.40 - 29.55 average years of the wake and sleep groups, respectively) is unlikely to account for the absence of differences in discriminability and reliability.



*Discriminability*

Figure 1A shows the ROC curves for the sleep and wake conditions. The points represent the data and the size of the points reflect the relative frequencies of each (Seale-Carlisle, Wetmore, Flowe & Mickes, 2019). All of the succeeding ROC and CAC figures feature relative frequencies. Figure 1A (and all ROC figures) are labelled with the overall filler ID rates from the target-absent lineups (on the top x-axis) and estimated false ID rates (bottom x-axis) to show that either way to analyse the ROC data is legitimate (Mickes et al., 2017). The curves in Figure 1A (and the ROC curves to follow) represent Ensemble unequal variance signal detection model fits (Wixted, Vul, Mickes & Wilson, 2018). Using a TA filler ID cut-off of .44, there was no difference between the pAUC values from the sleep (.28) and wake (.28) conditions,  $D = .14$ ,  $p = .892$ . As shown in Figure 1A by the size of the points, the most common suspect identification for the wake condition was made with 100% confidence and the most common suspect identification for the sleep condition was made with 90% confidence.

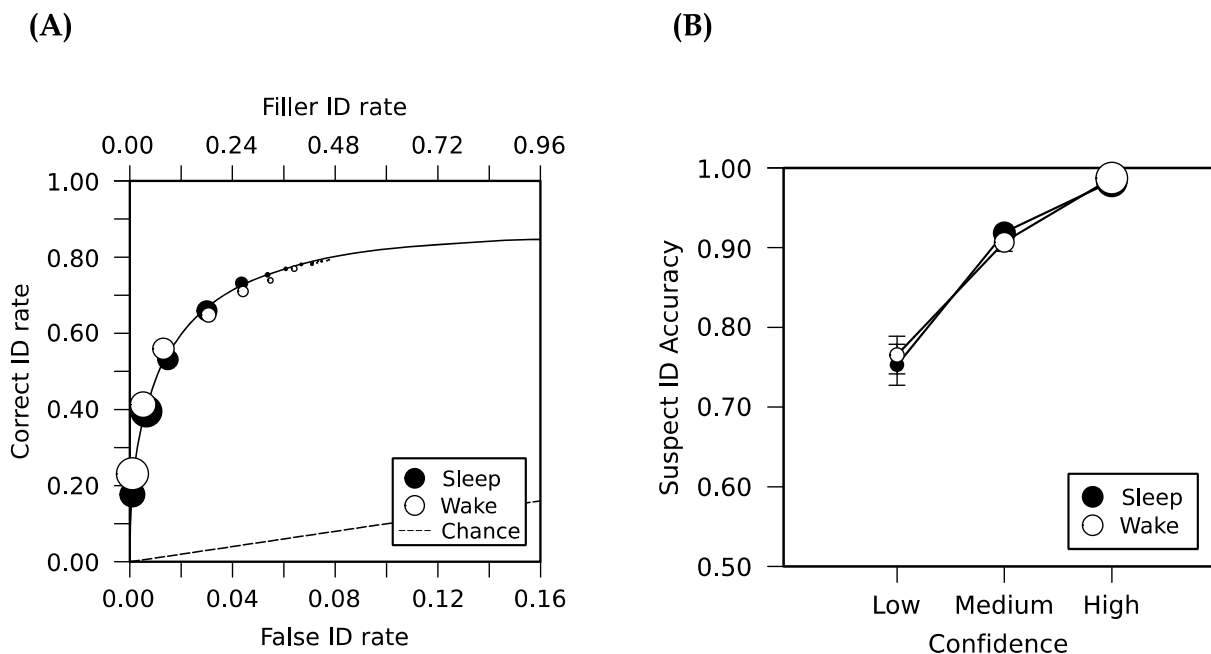


Figure 1 (A) ROC curves for sleep and wake groups. The points represent the data and the curves represent signal detection model fits. The overall filler ID rates from target-absent

lineups are shown on the top x-axis and the estimated false ID rates are shown on the bottom x-axis. **(B)** CACs for sleep and wake groups. The size of the points in A and B represent relative frequencies of responses and the bars represent standard errors.

### *Reliability*

Because there are few responses in some of the confidence bins, as is clear in the rightmost points in Figure 1A, we collapsed across confidence based on precedent (Mickes, 2015) in the following manner: 0-60%, 70-80%, and 90-100%. Figure 1B shows the confidence-accuracy characteristic (CAC) curves for the sleep and wake conditions. There were no differences in reliability and in both cases, the majority of responses were made with high confidence, followed by medium confidence, followed by low confidence.

### *Time-of-day Effects*

Table 3 shows the frequencies of all of the response types by level of confidence for target-present and target-absent lineups for the AM and PM conditions. Correct ID rates for the AM and PM conditions were both .87 and the estimated false ID rates were both .06. Figure 2A shows the ROC curves for the AM and PM conditions. Using a TA filler ID cut-off of .63, there was no difference between the pAUC values from the AM (.278) and PM (.271) conditions,  $D = .69$ ,  $p = .488$ . Likewise, as shown in Figure 2B, there were no differences between the AM and PM CAC curves. Thus, any differences that would have been found in the experimental conditions would not be explained by time of day effects. The AM and PM conditions yielded similar patterns of responses as the sleep and wake conditions, as reflected in the size the points in Figures 2A and 2B, whereby, in both conditions, suspects were identified most commonly with the highest level of confidence.

### *Current Sleepiness*

Across the sample, participants reported being sleepier in the evening compared to in the morning. The SSS scores were significantly higher for the sleep condition ( $M = 2.70$ ,  $sd = 1.19$ ) than the wake condition ( $M = 2.09$ ,  $sd = 1.01$ ) during the study phase (in the PM),  $t(1,998) = 12.27$ ,  $p < .001$ ,  $d = .55$ . The SSS scores were significantly higher for the wake condition ( $M = 2.72$ ,  $sd = 1.46$ ) than sleep condition ( $M = 2.28$ ,  $sd = 1.18$ ) during the test phase (in the PM),  $t(1,998) = 7.31$ ,  $p < .001$ ,  $d = .33$ . The same pattern arises with the control conditions. Because participants in the control conditions report SSS scores at the same time point (unlike those in the experimental conditions), we averaged the two SSS scores from each participant in the control conditions and those averages were significantly higher for the PM condition ( $M = 2.52$ ,  $sd = 1.19$ ) than the AM condition ( $M = 2.10$ ,  $sd = .99$ ),  $t(1,998) = 8.47$ ,  $p < .001$ ,  $d = .38$ .

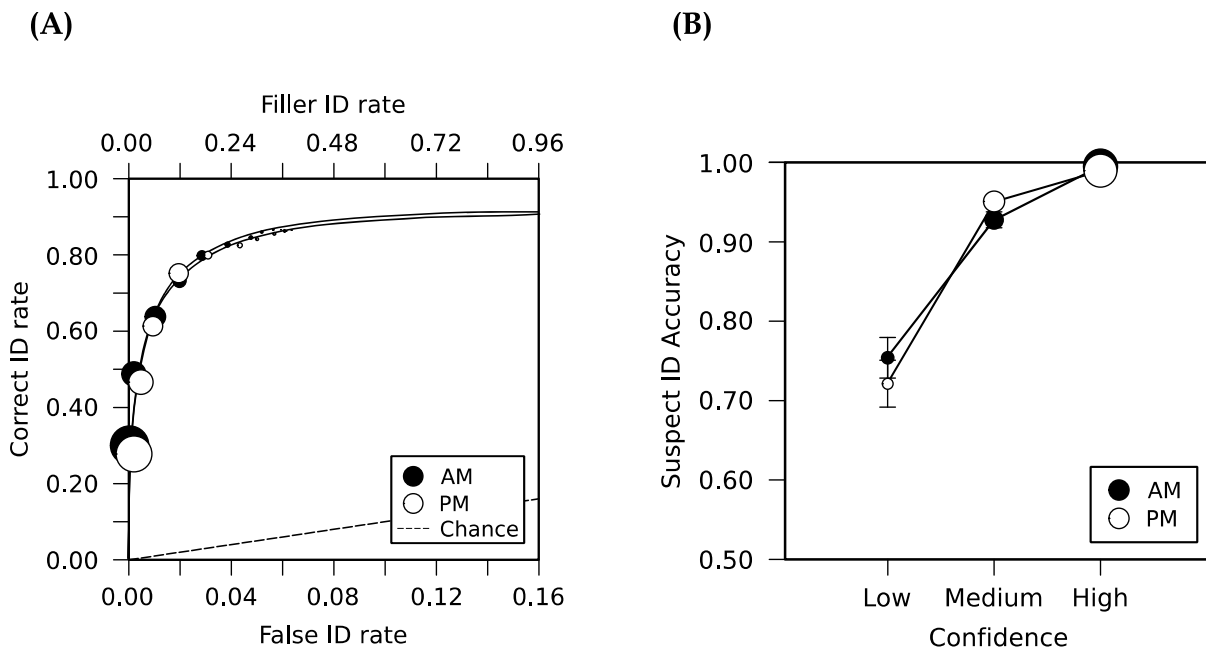
**Table 3.** Frequencies of correct IDs (CID), filler IDs (FID), and no-IDs for target-present and target-absent lineups for each level of confidence for sleep and wake groups.

Confidence	AM Control					PM Control				
	Target-present			Target-absent		Target-present			Target-absent	
	CID	FID	no ID	FID	no ID	CID	FID	no ID	FID	no ID
0	0	0	2	4	2	1	1	4	6	2
10	0	2	0	9	5	1	7	1	2	5
20	3	5	0	14	4	3	1	1	12	5
30	7	2	1	13	17	7	5	0	20	11
40	9	2	3	28	18	8	1	2	20	16
50	14	6	10	31	22	13	2	9	37	27
60	32	1	5	27	39	24	5	1	34	33
70	46	5	11	29	61	70	4	8	30	54
80	73	1	5	26	57	74	6	4	14	56
90	91	0	1	5	59	95	0	3	8	49
100	146	0	3	1	43	140	1	2	6	49
Total	486			514		504			496	

### *Exploratory Analyses*

To rule out any impact that caffeine may have on performance (Borota et al., 2014), participants in the sleep and wake conditions who reported consuming caffeine were removed.

We then conducted exploratory analyses on the remaining participants from the sleep ( $n = 918$ ) and wake ( $n = 887$ ) conditions. There was no discriminability difference between the pAUC of the sleep condition (.28) vs. the pAUC of the wake condition (.29),  $D = .53$ ,  $p = .593$  (a TA filler ID cut-off of .56 was used).



*Figure 2. (A)* ROC curves for AM and PM control groups. The operating points represent the data and the curves represent signal detection model fits. The overall filler ID rates from target-absent lineups are shown on the top x-axis and the estimated false ID rates are shown on the bottom x-axis. *(B)* CACs AM and PM control groups. The size of the points in A and B represent relative frequencies of responses and the bars represent standard errors.

There was also no reliability difference between the sleep and wake conditions. For the sleep condition, the proportion correct for identifications made with low, medium, and high confidence were 75%, 92%, and 98%, respectively. For the wake condition, the proportion correct identifications made with low, medium, and high confidence were 76%, 91%, and 99%, respectively. These results do not show support for changes in performance due to caffeine consumption.

To rule out any impact of poor sleep quality on performance, participants in the sleep condition who reported that they slept "very badly" ( $n = 3$ ), "badly" ( $n = 19$ ) or "fairly badly" ( $n = 113$ ) between study and test phases were removed. The remaining participants ( $n = 865$ ) had a mean SMHS score of 5.14, ( $sd = .52$ ,  $range = 4-6$ ). We then compared the discriminability of this subset of participants from the sleep condition with those in the wake condition. There was no difference between the pAUC of the sleep condition (.28) vs. the wake condition (.28),  $D = .19$ ,  $p = .845$  (a TA filler ID cut-off of .56 was used). There was also no difference in reliability between the sleep and wake conditions. For the sleep condition, the proportion correct for identifications made with low, medium, and high confidence was 74%, 92%, and 98%, respectively. For the wake condition, the proportion correct for identifications made with low, medium, and high confidence was 77%, 91%, and 99%, respectively. With the data from those who reported having poor sleep quality between the study and test phases removed, there were still no differences. The results in Figures 1A and 1B cannot be explained by poor sleep quality.

## Discussion

We conducted a large-scale investigation on the impact that sleep may have on eyewitness identification performance, namely discriminability and reliability. This is the largest sleep and episodic memory experiment to date. Based on prior literature, we did not have strong predictions about how sleep would impact reliability, but we did predict that sleep would benefit the ability to discriminate guilty suspects from innocent suspects. The results did not bear out that prediction.

### *Discriminability*

As predicted, there were no differences in discriminability between the time of day AM and PM control conditions. If there were differences between these control conditions, we

would not be able to make strong claims about any differences in the experimental conditions. The data did not, however, support the prediction that discriminability was greater for the sleep condition vs. the wake condition. Sleep did not improve memory. This result is counter to the many findings in the sleep literature (see reviews by Diekelmann & Born, 2010; Dudai, Karni & Born, 2015; Rasch & Born, 2013; Stickgold, 2005; Yonelinas, Ranganath, Ekstrom & Wiltgen, 2019). We consider and address several possible explanations for this discrepancy.

#### *Differences in sleepiness levels*

One possibility is that participants in the sleep condition were sleepier and therefore did not perform as well as they would have otherwise. We initially predicted no differences in sleepiness, based on the SSS scores, between the sleep and wake conditions or the AM and PM conditions (Maurer et al., 2015; Payne et al., 2012; Payne et al., 2008; Sheth et al., 2009), but there were differences. It was not the case, however, that the sleep condition uniformly reported being sleepier. Participants in all conditions reported being sleepier when asked in the evening. Participants in the sleep condition encoded the information in the evening, when they reported being sleepier. If being sleepier at encoding affected discriminability, then the sleep condition may have had greater discriminability if they were not sleepy during the study phase.

Participants in the PM condition also reported being sleepier during the study phase, but there were no differences between the control conditions. It is therefore unlikely that the reason the sleep condition did not have greater discriminability than the wake condition is because they were sleepier during encoding. One could argue, however, that based on the encoding specificity principle (Tulving & Thomson, 1971), the PM controls, even though they were sleepy at encoding, did better because they were sleepy both at encoding and at test (i.e., same state of sleepiness), and for this reason they performed as well as their AM counterparts.

In comparison the state of sleepiness was not consistent for the participants in the experimental conditions between study and test. To address this possible concern, we selected an equal number of participants from sleep and wake conditions who responded with a 1, 2, or 3 on the SSS on both occasions. By doing this, we screened out the participants who reported being sleepy (i.e., responded with a 4, 5, 6, or 7 on the SSS on both occasions), in the sleep ( $n = 247$ ;  $M = 1.87$ ,  $sd = 0.88$ ) and wake ( $n = 247$ ;  $M = 1.87$ ,  $sd = 0.88$ ) conditions and then compared discriminability. Though the sample sizes were reduced, we still had 95% power to detect a medium sized effect, one that is comparable in the sleep research literature (Gui et al., 2017). The pAUC for the sleep condition (.23) was not significantly greater than the pAUC for the wake condition (.20),  $D = 1.32$ ,  $p = .188$  (with the TA filler ID pAUC cut-off of .62). Therefore, two lines of evidence suggest that the higher sleepiness does not explain the absence of a sleep benefit. First, participants in the PM condition were sleepier during the study phase and this did not negatively impact their discriminability when compared to the participants in the AM control condition. Second, matching the levels of sleepiness between sleep and wake conditions during the study and test phases and comparing discriminability yielded no overall benefit of sleep on memory.

#### *Differences in types of tests*

Another possible reason that the sleep condition did not outperform the wake condition in the current experiment may be because the sleep benefit is more pronounced when memory is tested on a recall test and less pronounced when memory is tested on a recognition test (Diekelmann, Wilhelm & Born, 2009). However, in a recent meta-analysis of over 40 published experiments, the effect sizes were medium when memory was tested on cued recall, free recall, and recognition memory tests (.55, .49, and .44, respectively) (Berres & Erdfelder, 2018). We should therefore be able to detect a sleep benefit even when using a recognition memory test. After all, our experiment has the power to detect a small effect.

Maybe it is just not the right type of recognition memory test. There are similarities and differences between standard list-learning and forensically-relevant experiments. One similarity is that they both have target-absent and target-present trials. However, in list-learning experiments, targets and lures are shown one at a time and a decision is made on each item. When memory is tested on a lineup, the items are presented simultaneously and one decision is made. Another difference is the number of trials. There are multiple trials per participant in list-learning experiments and only one trial per participant in forensically-relevant experiments. Therefore, these types of recognition memory experiments differ in the relative influences of the sources of within- and between-subjects variance (Benjamin, Diaz & Wee, 2009; Wixted et al., 2018). If the memory benefits afforded by sleep are conceptually replicable, then these differences should not account for the fact that in some of the list-learning experiments sleep enhanced memory whereas there was no benefit of sleep in our experiment. Thus, differences in how memory is probed, either in a free recall, cued recall, or a type of recognition memory test, is an insufficient explanation for the lack of a sleep benefit in our experiment.

*Comparing one forensically-relevant experiment with another*

Although we did not find an advantage of sleep on discriminability in a forensically-relevant experiment, a recent paper reported one. In our introduction, we wrote that researchers had not yet investigated the impact of sleep on eyewitness identifications. However, this changed while our data collection was underway. A paper on the topic was published in which the findings of two forensically-relevant experiments using an AM-PM:PM-AM design were reported (Stepan, Dehnke & Fenn, 2017). In both experiments, participants watched a video of a mock crime, remained awake during the day or slept overnight, and then memory for the perpetrator was tested on a lineup. In one experiment, participants were only tested on target-present lineups and in the other experiment, participants were only tested on target-absent



lineups. In the former, there were no effects of sleep vs. wake on correct identifications, filler identifications, or misses. In the latter, there were effects of sleep vs. wake on false identifications but not on correct rejections. There were fewer false identifications in the sleep condition.

To measure discriminability, both target-present and target-absent lineups need to be included in the same experiment (Mickes, 2016; Mickes & Wixted, 2015; Rotello & Chen, 2016). Measuring correct ID rates and false ID rates separately has the potential to mislead (Clark, 2012). Lower false ID rates can indicate more conservative responding or increased discriminability. If the former, the lower false ID rate would be accompanied by a lower correct ID rate. This would not mean that memory is better. Instead it would mean that one condition is less likely to make an identification, so fewer innocent suspects would be identified but fewer guilty suspects would be identified as well. That is the reason why participants should be tested on both target-present and target-absent lineups in one experiment and why correct IDs and false IDs need to be considered together (Mickes, 2016; Mickes & Wixted, 2015).

To assess whether we replicated the findings reported in Stepan et al. (2017) we conducted the same analyses on the responses from the target-present lineups (a 2 (sleep, wake) x 3 (correct IDs, filler IDs, misses) chi-square analysis) and then on the responses from target-absent lineups (a 2 (sleep, wake) x 2 (filler IDs, correct rejections) chi-square analysis). To match their sample sizes, we took the first 88 participants assigned to target-present lineups in the wake ( $n = 47$ ) and sleep ( $n = 41$ ) conditions and the first 96 participants assigned to target-absent lineups in the wake ( $n = 53$ ) and sleep ( $n = 43$ ) conditions. There were no differences between conditions in response types from the target-present lineups,  $\chi^2(2, N = 88) = 1.83, p = .759$ . There were also no differences between conditions in response types from the target-absent lineups,  $\chi^2(1, N = 96) = .014, p = .904$ . Thus, the results from the target-present lineups replicated those reported in Stepan et al., but the results from the target-absent lineups did not.

The fact that each participant provides only one data point means the sample size needs to be sufficiently large to accommodate for this. Because there were fewer than five responses in some of the cells, and chi-square analysis is sensitive to small sample sizes, we conducted the same analysis on our entire sample (Bewick, Cheek & Ball, 2003; McHugh, 2013). Again, there were no differences between conditions in response types from target-present lineups,  $\chi^2(2, N = 2000) = .321, p = .571$ , or target-absent lineups,  $\chi^2(1, N = 2000) = 0.65, p = .420$ . These results map onto the results from ROC analysis (a more suitable analysis than separate chi-square analyses) in which there were no differences between the experimental conditions.

Why did Stepan et al. (2017) find a decreased false ID rate in the sleep condition compared to the wake condition and we did not? Our experiments were similar in that we used the same AM-PM:PM-AM design, memory was tested on a lineup, and participants provided only one response. Despite these similarities, there were differences: we 1) did not use the same stimuli or procedures, 2) did not conduct separate experiments for target-absent and target-present lineups, 3) conducted the experiment online, and 4) had a sample size that was approximately 10 times larger. Therefore, maybe a direct replication would find the same results. Another possible reason for the discrepancy between our experiments is that the lower false ID rate in the sleep condition in Stepan et al. was a false positive result (Simmons et al., 2011).

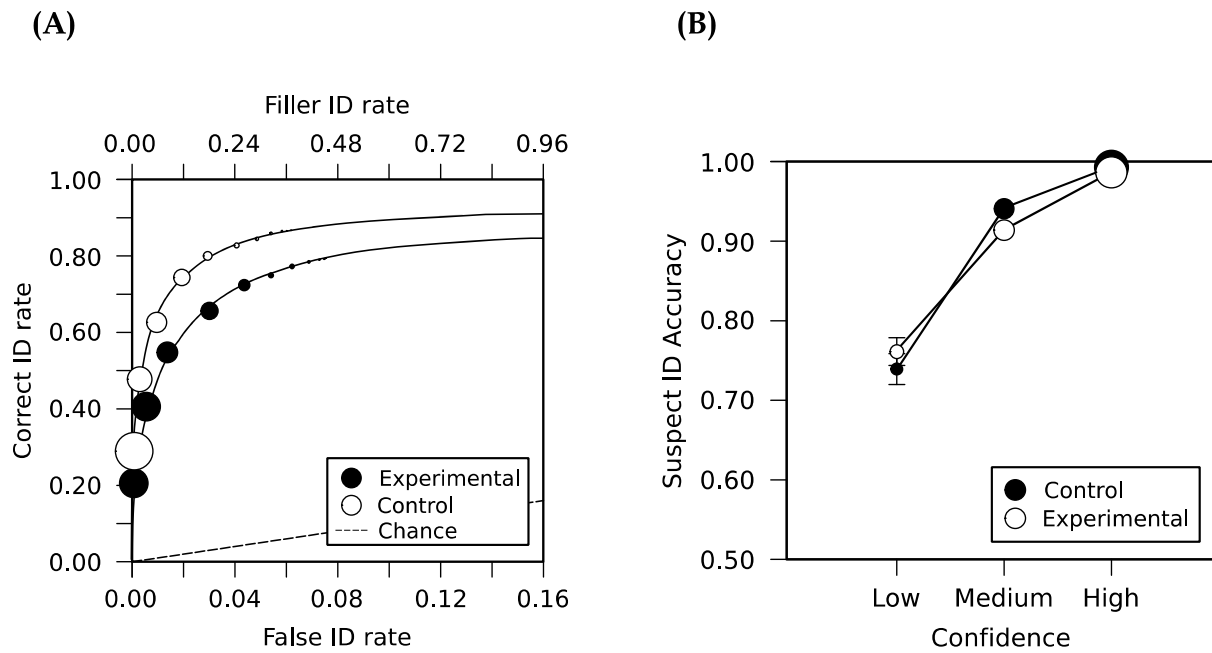
### *Reliability*

We did not make strong predictions about any effects that sleep may have on reliability (i.e., the likelihood that the suspect identified was indeed guilty). The sleep and wake conditions did not differ in reliability. Overall, the confidence that participants expressed was informative of accuracy. That is, averaged across all conditions, high confidence identifications

were higher in accuracy than medium confidence identifications, which were higher than accuracy than low confidence identifications (99% vs. 93% vs. 75%, respectively).

One nuanced, and important, concept is that even if discriminability is low, reliability can be high (Palmer et al., 2013; Semmler, Dunn, Mickes & Wixted, 2018). Discriminability was lower for our experimental conditions compared to our control conditions. If this were not so, then we would have detected a problem because the experimental conditions were tested after a 12-hour retention interval and the control conditions were tested after a 5-minute retention interval. We conducted an exploratory analysis to determine whether reliability would be the same despite discriminability differing. Because there were no differences between the sleep and wake conditions and the AM and PM conditions, we collapsed across and compared the pAUC values of the experimental conditions with the control conditions. As shown in Figure 3A, the pAUC for the control condition (.276) was significantly greater than the pAUC for the experimental condition (.229),  $D = 5.92$ ,  $p < .001$  (with the false ID pAUC cut-off of .63).

Despite differences in discriminability between the experimental and control conditions, identifications made with high confidence were highly accurate (averaged across all conditions, they were 99% accurate). Figure 3B shows the CACs curves for the experimental and the control conditions. Thus, even though discriminability was lower for the experimental conditions than the control conditions (because there was a 12-hour retention interval vs. a 5-minute retention interval) participants were similarly well-calibrated. That is, in the experimental and control conditions, identifications made with high confidence are higher in accuracy than identifications made with medium confidence which are higher in accuracy than identifications made with low confidence (and the standard error bars are overlapping). This finding joins other findings in which even for lower discriminability conditions, reliability remains strong (Semmler et al., 2018; Wixted & Wells, 2017).



*Figure 3. (A)* ROC curves for experimental and control groups. The operating points represent the data and the curves represent signal detection model fits. The overall filler ID rates from target-absent lineups are shown on the top x-axis and the estimated false ID rates are shown on the bottom x-axis. *(B)* CACs for experimental and control groups. The size of the points in A and B represent relative frequencies of responses and the bars represent standard errors.

### *Conclusion*

Sleep joins the growing list of variables that do not impact experimental eyewitnesses' reliability (Mickes, 2015; Wixted & Wells, 2017). It also did not affect discriminability. Thus, the widely touted benefits of sleep were not found in a large-scale forensically-relevant eyewitness identification experiment. Ours is not the first published experiment to find evidence that differs from most reports in the sleep research literature (Hallgato, Gyori-Dani, Pekar, Janaxsek & Nemeth, 2013; Rickard, Cai, Jones & Ard, 2008). Likewise, we failed to conceptually replicate the general findings reported in the literature that sleep benefits

recognition memory. These findings challenge the widely claimed advantage that sleep has on recognition memory over a 12-hour period. Further well-powered conceptual and direct replications should clarify the benefits of sleep on recognition memory.

## Chapter 3: Norms for Sleepiness Scales and Sleep Duration in a Young Adult American Sample

### Abstract

Sleep impacts various aspects of health and cognitive functioning. To understand the relationship between poor sleep and health, researchers and clinicians use self-report tools such as the Stanford Sleepiness Scale (SSS) and Epworth Sleepiness Scale (ESS) and measure sleep duration. To make reliable comparisons between clinical and healthy populations, clinicians and researchers use normative values for these outcomes. To date, no norms exist for the SSS and there are limited data on young adults for the ESS. We aimed to provide norms for the SSS, ESS, and sleep duration and determine whether they vary based on time of day, age, education and sex. This was a survey study; data were collected on the SSS and ESS to measure sleepiness and sleep duration was measured by asking the number of hours and minutes slept. Data were collected online in the United States as part of a large-scale experiment investigating sleep and memory (Morgan et al., 2019). In that experiment participants provided demographic information, sleep duration for the night before participation, completed the SSS, took part in a memory task, and completed the ESS. Participants were aged between 18-40-years. Exclusion criteria included current diagnosis of a sleep, neurological or psychiatric disorder, currently under medication affecting sleep or memory, shift work, and travel across time zones in the past two weeks. There were 7533 eligible individuals who met the criteria and participated (age  $M = 29.77$  years,  $SD = 5.60$ ; Males = 3035, Females = 4468). SSS scores were greater in the evening ( $\tilde{x} = 3$ ,  $Q1 = 2$ ,  $Q2 = 3$ ) than the morning ( $\tilde{x} = 2$ ,  $Q1 = 2$ ,  $Q2 = 3$ ),  $p < .001$ ,  $n_p^2 = .044$ . Other statistically significant differences of age, education, and time of day were found for the SSS and ESS but the effect sizes were trivially small. Sleep duration varied by age, but this effect was also very small. SSS scores vary by time of day, therefore clinicians and researchers

should consider when the scale is administered in determining the sleep health of patients and participants. The normative data can be used by clinicians and researchers.

## Introduction

Sleep is vital for maintaining physical and mental health (e.g., Garbarino, Lanteri, Durando, Magnavita & Sannita, 2016; Scott, Webb & Rowse, 2017). Sleep quality ranges from very good to very bad; on the poorer end, lack of sleep and sleep disruption can affect every day functioning, including cognitive functioning. There are over 100 sleep disorders (Kumar, 2008) and sleep related problems are associated with the onset, development, and maintenance of mental health disorders (e.g., Baglioni et al., 2011; Cox & Olatunji, 2016). As such, sleep abnormalities are criteria for some mental disorders. For example, too many or too few hours of sleep are criteria for some depressive disorders (DSM-V, 2013). Problems with sleep co-occur with numerous diseases including cardiovascular disease (Bauters, Rietzschel, Hertegonne & Chirinos, 2016), pathologies that lead to stroke (Hepburn, Bollu, French & Sahota, 2018; Lau, Rundek & Ramos, 2019), Alzheimer's disease and other dementia-causing diseases (Irwin & Vitiello, 2019), and diabetes (Shan et al., 2015). Problems with sleep can also lead to impaired cognitive functioning (e.g., poor working memory; Fortier-Brochu, Beaulieu-Bonneau, Ivers & Morin, 2012).

Sleep difficulties negatively affect people with disorders and diseases and also people without, which in turn can compromise day-to-day performance. Sleep related problems in the general population are increasing in prevalence. Physician visits in the United States where patients report sleep disturbances went from approximately 6 million in 1999 to approximately 8 million as of 2010 (Ford et al., 2014). In response, reports related to sleep and health have been increasing too (Center for Disease Control, 2005; 2008; 2009; Altevogt, & Colten, 2006; Hafner, Stepanek, Taylor, Troxel & van Stolk, 2017; Hirshkowitz et al., 2017; National Sleep Foundation, 2013; Liu et al., 2016; Ohayon et al., 2017; Watson et al., 2015). The United States Department of Health and Human Services (2015) launched the Healthy People 2020 campaign that is aimed at improving sleep. And the estimated cost of sleep related problems has risen



from hundreds of millions of dollars to approximately \$400 billion as of 2017 in the United States (Hafner et al., 2017). These costs include expenditures that are associated with the reduction in productivity and time lost working (Hafner et al., 2017; Skaer & Sclar, 2010).

Sleep-related scales are used to better understand sleep and sleep-related issues by clinicians and researchers. While some are widely used, many of the scales do not have any, or have limited, normative data. We provide normative data of a large sample of young American adults by age, sex, education, and time of day for two scales of sleepiness, the Stanford Sleepiness Scale (SSS, Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973) and the Epworth Sleepiness Scale (ESS, Johns, 1991), and sleep duration. Both scales are described in detail below.

#### *Stanford Sleepiness Scale (SSS) Background and Use*

The SSS (Hoddes et al., 1973) measures the current level of sleepiness and is commonly used by researchers (cited 2301 times, Google Scholar, 2019). One example of its use in research is to measure increases in sleepiness when participants are sleep deprived (e.g., Goldich et al., 2010; Rupp, Killgore, & Balkin, 2010; Urrila, Stenuit, Huhdankowski, Kerkhofs & Porkka-Heiskanen, 2007). Researchers have correlated performance on various cognitive tasks with SSS scores. There is evidence that increases in sleepiness are related to decreases in cognitive performance, such as with attention (e.g., Goldich et al., 2010). In addition to capturing the relationship between sleepiness and behavioural outcomes, the SSS has been used to rule out sleepiness as a confounding variable (e.g., Maurer et al., 2015; Morgan et al., 2019; Payne et al., 2012; Payne, Stickgold, Swanberg & Kessinger, 2008; Sheth, Nguyen & Janvelyan, 2009). For example, to rule out the possibility that the level of sleepiness affects cognitive performance, participants who are sleepy, based on their SSS scores, would be excluded from the analyses. Instead of eliminating data from participants who scored above a certain level by using an arbitrary cut-off, a comparison of the scores to normative data could

be used. That way, individual scores that are outliers based on age and sex can be removed and possibly fewer participants would have to be excluded from the analyses.

The SSS is underused in clinical environments, in large part because of the lack of normative data minimizes its clinical usefulness (Hirshkowitz, Sarwar, & Sharafkhaneh, 2011). By providing normative SSS data clinicians will be able to make relevant comparisons between their patients' scores and scores from age and sex-matched participants, thereby aiding their decision-making about sleep health.

#### *Epworth Sleepiness Scale Background and Use*

While the SSS measures the current state of sleepiness (Hoddes et al., 1973) the ESS measures general daytime sleepiness (Johns, 1990). The ESS is one of the most widely used sleepiness scales; to date it has been cited over 12,000 times, which has doubled in just seven years (Google Scholar, 2019). The ESS has been used to measure daytime sleepiness in the general population (e.g., Beaudreau et al., 2012; Gander et al., 2005; Joo et al., 2005; Sauter et al., 2007; Spira et al., 2012; Wu et al., 2012) and clinical populations, in patients with neurological disorders (e.g., Chung et al., 2013; De Cock et al., 2014; Rammohan et al., 2002; Rosenthal & Dolan, 2008; Teodorescu et al., 2006) and sleep disorders (e.g., Kloepper et al., 2009). The ESS has also been evaluated for use as a diagnostic tool to measure the likelihood of having a sleep disorder, such as narcolepsy (Nishiyama et al., 2014).

In terms of the use of the ESS in the lab, it has been used in ways similar to the SSS. For example, the relationship between daytime sleepiness and sustained attention has been measured (e.g., Fronczek, Middelkoop, & van Dijk & Lammers, 2005). Also, data from participants have been excluded on the basis of ESS scores that reflect excessive sleepiness (e.g., Ellenbogen, Hulbert, Stickgold, Dinges, & Thompson-Schill, 2006). The ESS has also been used to rule out daytime sleepiness as a potential confound (e.g., Wiebe, Carrier, Frenette & Gruber, 2013).

Although there currently are normative ESS data (Sander et al., 2016; Sanford et al., 2006; Sauter et al. 2007; Spira et al., 2011; Stavem et al., 2017) data from young adults aged between 18-40 years are largely underrepresented. Across the normative studies the mean age is 61 years. In just the way normative SSS data can be used by researchers and clinicians, so too can normative ESS data. We provide data from a large young adult sample.

#### *Sleep Duration Background and Use*

Sleep duration is used as an obvious indicator of poor sleep and sleepiness (e.g., Steptoe, Peacey, & Wardle, 2006). Clinically short and long sleep duration are associated with various disorders which impact everyday cognitive functioning and health (Groeger, Zijlstra & Dijk, 2004; Stamatakis, Kaplan & Roberts, 2007; Steptoe et al., 2006). Norms of sleep duration can be used clinically to determine whether a patient's sleep duration significantly deviates from the norm, and in turn whether they are at risk of certain disorders. In the lab sleep duration is measured to ensure that participants have slept, and are often arbitrarily excluded from analyses if they have slept less than 6 hours (e.g., Sheth, Nguyen & Janvelyan, 2009; Morgan et al., 2019). Ideally sleep duration measured in the lab should be compared to norms of sleep duration to determine a participant's eligibility for inclusion in data analysis.

There are normative sleep duration data of American adolescents and young adults aged 12-32 years (e.g., Maslowsky & Ozer, 2014). Our sample therefore overlaps in age by 14 years (ages 18-32 years). Possible explanations for the interesting differences between the two samples emerged. Whatever the reasons, a more modern dataset seems warranted.

#### *Scope and aims of the current study*

We provide normative data of a young (18-40 years of age) adult American population for the SSS, ESS, and sleep duration. The data reported here were collected as part of an

experiment in which participants completed the SSS and ESS, provided the length of time that they slept the night before, and took part in a memory experiment in the morning between 8am-11am and/or in the evening between 8pm-11pm (Morgan et al., 2019). SSS and ESS scores were compared across age, education, and sex. Because participants took the SSS and ESS in either the morning or the evening, those scores could be compared. Sleep duration was compared across age, education and sex. The normative data provided here can be used by researchers investigating sleep-related issues and clinicians make decisions about the sleep health of individuals.

## Methods

### *Participants*

Participants ( $N = 7533$ ) from the United States across the Central, Eastern, Mountain and Pacific time zones were recruited using the work source site, Amazon Mechanical Turk (<https://www.mturk.com/>). Table 4 shows age ( $M = 29.77$  years,  $SD = 5.60$ ), ethnicity, and level of education by sex for the entire sample. The entire sample completed both the SSS and provided the previous night's sleep duration. The average hours slept was 7.59 ( $SD = 3.16$ , range = 0-15 hours). Table 5 which shows age ( $M = 29.94$  years,  $SD = 5.53$ ), ethnicity, and level of education by sex for the subset of the participants ( $n = 4355$ ) who completed the SSS, provided the previous night's sleep duration and also completed the ESS.

The inclusion criteria were to be between 18-40 years of age. The exclusion criteria were 1) a current diagnosis of any sleep disorder, psychiatric disorder or neurological disorder that affects memory; 2) currently engaged in shift work; 3) have travelled across time zones in the past two weeks; and 4) be currently taking any prescribed medication that may affect sleep or memory. Ethical approval was granted by the Royal Holloway, University of London Ethics Committee; project number 295-2019-08-27-16-20.

**Table 4.** Demographic data for all participants.

	Males ( <i>n</i> = 3035)		Females ( <i>n</i> = 4468)		Total ( <i>N</i> = 7533)*
	<i>n</i>	(%)	<i>n</i>	(%)	<i>n</i>
<b>Age Group</b>					
18-25	882	(29.0)	1035	(23.2)	1930
26-30	873	(28.7)	1271	(28.4)	2151
31-35	817	(26.9)	1247	(27.9)	2076
36-40	463	(15.3)	915	(20.5)	1384
<b>Ethnicity</b>					
African American	200	(6.6)	453	(10.1)	654
Arab American	22	(0.7)	21	(0.5)	43
Asian American	292	(9.6)	278	(6.2)	573
Caucasian American	2124	(69.9)	3137	(70.2)	5270
Hispanic American	204	(6.7)	288	(6.4)	493
Indian American	59	(1.9)	76	(1.7)	138
Native American	12	(0.4)	23	(0.5)	35
Other	82	(2.6)	149	(3.3)	231
Do not wish to state	40	(1.3)	43	(0.9)	99
<b>Education</b>					
At least bachelor's degree	2171	(71.5)	2726	(61.0)	4897
Less than bachelor's degree	1149	(37.8)	1732	(38.7)	2881
Do not wish to state	6	(0.2)	10	(0.2)	16

\*Excludes participants who responded "other" (*n* = 2) and "do not wish to state" (*n* = 28)

### Materials

*Stanford Sleepiness Scale (SSS)*. The SSS is composed of eight options from which participants choose which best suits their current state (Hoddes et al., 1973). The options are: 1) feeling active, vital, alert or wide awake, 2) functioning at high levels, but not at peak; able to concentrate; 3) awake, but relaxed; responsive but not fully alert; 4) somewhat foggy, let down; 5) foggy; losing interest in remaining awake; slowed down; 6) sleepy, woozy, fighting sleep; prefer to lie down; 7) no longer fighting sleep, sleep onset soon; having dream-like thoughts; and 8) asleep. The numbers correspond with the points assigned to the option. For example, a participant's response of "somewhat foggy, let down" would be given a score of 4. Higher scores indicate higher levels of current sleepiness.

**Table 5.** Demographic data of the subset who completed the SSS and ESS.

	Males ( <i>n</i> = 1720)		Females ( <i>n</i> = 2620)		Total ( <i>N</i> = 4355)*
	<i>n</i>	(%)	<i>n</i>	(%)	<i>n</i>
<b>Age Group</b>					
18-25	477	(28.0)	568	(22.0)	1050
26-30	483	(28.0)	745	(28.4)	1228
31-35	494	(28.0)	767	(29.3)	1269
36-40	266	(15.0)	540	(20.6)	808
<b>Ethnicity</b>					
African American	102	(5.9)	252	(9.6)	354
Arab American	12	(0.7)	9	(0.3)	21
Asian American	155	(9.0)	147	(5.6)	304
Caucasian American	1245	(72.4)	1930	(73.7)	3180
Hispanic American	110	(16.4)	136	(5.2)	247
Indian American	30	(1.7)	40	(1.5)	71
Native American	3	(0.2)	12	(0.5)	15
Other	37	(2.2)	74	(2.8)	111
Do not wish to state	26	(1.5)	20	(0.8)	52
<b>Education</b>					
At least bachelor's degree	1094	(63.6)	1665	(63.5)	2759
Less than bachelor's degree	622	(36.2)	949	(36.2)	1571
Do not wish to state	4	(0.2)	6	(0.2)	10

\*Excludes participants who responded "other" (*n* = 1) and "do not wish to state" (*n* = 14)

*Epworth Sleepiness Scale (ESS).* On the ESS, participants rate the likelihood of dozing (would never doze, slight chance, moderate chance, and high chance of dozing) in eight everyday scenarios (John, 1991). Responses to each item is scored from 0-3 yielding a total possible score of 24. Higher scores indicate higher levels of daytime sleepiness; scores greater than 10 are considered indicative of excessive daytime sleepiness (John, 1991).

*Sleep Duration.* Participants were asked "How long did you sleep last night?" and provided their responses in hours and minutes. The response options for hours ranged from 0-24 increasing by 1-hour intervals, and the options for minutes ranged from 0-55, increasing by 5-minute intervals.

### *Procedure*

The data were collected in an experiment in which the question of interest regarded the effect of sleep on eyewitness memory (Morgan et al., 2019). In that experiment, participants

were randomly allocated to one of four different conditions depending on the time of day that they chose to take part at: Sleep, wake, AM Control and PM control. Participants in the AM control condition completed the experiment in the morning and participants in the PM control condition completed the experiment in the evening. Those in the sleep condition completed the first part of the experiment in the evening and the second part in the morning. Lastly participants in the wake condition completed the first part of the experiment in the morning and completed the second part in the evening. Participation took place online. Participants consented, provided demographic information, information about sleep duration, took the SSS, took part in the memory task, took the ESS, and were debriefed.

### *Statistical analyses*

A correlation analysis was conducted on number of hours slept and the SSS and ESS scores. Because there were few participants of some ages the data were binned into the following age groups: 18-25, 26-30, 31-35, and 36-40 years. For the same reason, education was also binned into two groups: less than bachelor's degree and at least bachelor's degree. A 2 (sex) x 2 (education) x 4 (age group) between-subjects ANOVA was conducted to measure differences in sleep duration. To measure differences in sleepiness, the dependent variables were the SSS and ESS scores. The independent variables were age group, sex (male vs. female), and time of day (morning vs. evening). Two separate 2 (sex) x 2 (time of day) x 2 (education) x 4 (age group) between-subjects ANOVAs were conducted on the SSS and ESS scores. Interactions were assessed only if previous research suggested there were interactions. The raw data and analysis script are available on <https://osf.io/hnp5z/>.

## Results

### *Stanford Sleepiness Scale: Effects of Age, Time of Day, Education, and Sex*

Figure 4 shows means SSS scores provided in the morning and evening by age group and education level for females (Figure 4A) and males (Figure 4B). Table 6 shows the SSS scores that were provided in the morning or evening, and categorised by age, time of day, education, and sex. The average SSS score was 2.38 ( $SD = 1.18$ ). There was a negative correlation between the SSS scores and sleep duration ( $r = -.166, p < .001$ ), thus as sleep duration decreases scores on the SSS increase. There was a significant decrease in sleepiness with age,  $F(3,7467) = 18.06, p < .001, \eta_p^2 = .007, 90\% \text{ CI } [.004, .010]$ . SSS scores were also significantly greater in the evening compared to the morning,  $F(3,7467) = 342.84, p < .001, \eta_p^2 = .044, 90\% \text{ CI } [.037, .052]$ .

There was a significant effect of education: SSS scores were greater for the at least bachelor's degree group compared to the less than bachelor's degree group  $F(1, 7467) = 7.27, p = .007, \eta_p^2 = .001, 90\% \text{ CI } [.000, .003]$ . There was no significant effect of sex on sleepiness,  $F(1, 7467) = 2.75, p = .097, \eta_p^2 < .001, 90\% \text{ CI } [.000, .001]$ , and there was no interaction between sex and age,  $F(3, 7467) = 1.68, p = .169, \eta_p^2 = .001, 90\% \text{ CI } [.000, .002]$ .

Pairwise comparisons of SSS scores by age were made using Bonferroni corrections. The significant differences ( $p < .001$ ) were between 18-25-year olds and 31-35-year-olds, between 26-30-year olds and 31-35-year olds, 26-30-year olds and 36-40-year olds, and between 18-25-year olds and 36-40-year olds.



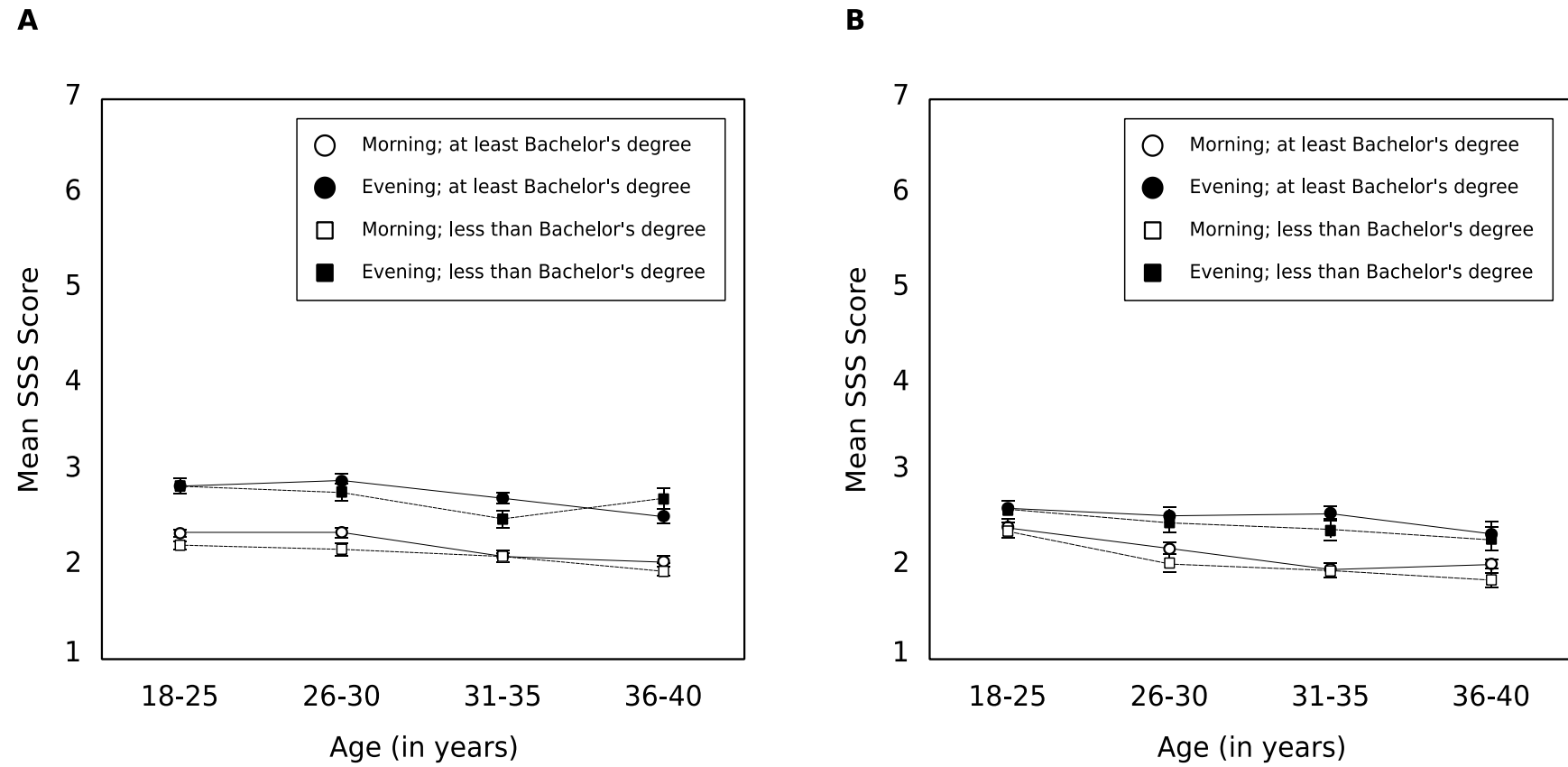


Figure 4. A) Average SSS scores for females by time of day and education. B) Average SSS scores for males by time of day and education.

Error bars represent standard error.

**Table 6.** SSS scores split by time of day, education, gender, and age category.


Morning																														
Age (Yrs)	Less than college educated															At least College Educated														
	Male					Female					Total					Male					Female					Total				
	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3
18-25	2.37	1.18	2.00	1.00	3.00	2.22	1.06	2.00	1.00	3.00	2.29	1.12	2.00	1.00	3.00	2.39	1.17	2.00	1.00	3.00	2.31	0.98	2.00	2.00	3.00	2.34	1.06	2.00	2.00	3.00
26-30	2.02	1.04	2.00	1.00	3.00	2.15	1.10	2.00	1.00	3.00	2.10	1.08	2.00	1.00	3.00	2.17	0.99	2.00	1.00	3.00	2.34	1.13	2.00	1.00	3.00	2.26	1.07	2.00	1.00	3.00
31-35	1.94	1.02	2.00	1.00	3.00	2.09	1.07	2.00	1.00	3.00	2.04	1.06	2.00	1.00	3.00	1.96	1.02	2.00	1.00	3.00	2.07	1.08	2.00	1.00	3.00	2.02	1.05	2.00	1.00	3.00
36-40	1.85	0.95	2.00	1.00	3.00	1.93	1.04	2.00	1.00	3.00	1.95	1.01	2.00	1.00	3.00	1.98	1.04	2.00	1.00	3.00	2.02	1.07	2.00	1.00	3.00	2.01	1.06	2.00	1.00	3.00
Total	2.09	1.09	2.00	1.00	3.00	2.11	1.07	2.00	1.00	3.00	2.10	1.08	2.00	1.00	3.00	2.11	1.05	2.00	1.00	3.00	2.18	1.08	2.00	1.00	3.00	2.15	1.07	2.00	1.00	3.00
Evening																														
Age (Yrs)	Less than college educated															At least College Educated														
	Male					Female					Total					Male					Female					Total				
	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3
18-25	2.59	1.17	3.00	2.00	3.00	2.85	1.28	3.00	2.00	3.00	2.72	1.24	3.00	2.00	3.00	2.62	1.18	3.00	2.00	3.00	2.83	1.20	3.00	2.00	3.00	2.73	1.19	3.00	2.00	3.00
26-30	2.44	1.22	3.00	1.00	3.00	2.77	1.25	3.00	2.00	3.00	2.64	1.25	3.00	2.00	3.00	2.55	1.22	3.00	2.00	3.00	2.92	1.24	3.00	2.00	3.00	2.77	1.24	3.00	2.00	3.00
31-35	2.37	1.19	2.00	1.00	3.00	2.48	1.30	2.00	1.00	3.00	2.44	1.26	2.00	1.00	3.00	2.54	1.23	3.00	1.00	3.00	2.71	1.26	3.00	2.00	3.00	2.63	1.26	2.00	1.00	3.00
36-40	2.27	0.99	2.00	2.00	3.00	2.69	1.36	3.00	2.00	3.00	2.56	1.27	3.00	2.00	3.00	2.34	1.27	2.00	1.00	3.00	2.51	1.19	3.00	1.00	3.00	2.45	1.22	3.00	2.00	3.00
Total	2.48	1.17	3.00	1.00	3.00	2.72	1.29	3.00	2.00	3.00	2.62	1.25	3.00	2.00	3.00	2.53	1.22	3.00	1.00	3.00	2.75	1.24	3.00	2.00	3.00	2.66	1.23	3.00	2.00	3.00

*Epworth Sleepiness Scale: Effects of Age, Time of Day, Education, and Sex*

Figure 5 shows means ESS scores provided in the morning and evening by age group and education level for females (Figure 5A) and males (Figure 5B). The average ESS score was 6.92 ( $SD = 3.18$ ). There was a significant negative relationship between ESS scores and sleep duration ( $r = -.071$ ,  $p < .001$ ): as sleep duration decreases scores on the ESS increase. Table 7 shows the breakdown of ESS normative scores across age, time of day, education, and sex. Figure 5 shows that ESS scores decreased with age,  $F(3,4320) = 7.96$ ,  $p < .001$ ,  $n_p^2 = .005$ , 90% CI [.002, .009]. ESS scores were higher in the evening compared to the morning,  $F(1, 4320) = 14.62$ ,  $p < .001$ ,  $n_p^2 = .003$ , 90% CI [.001, .007].

There was a significant effect of education: ESS scores were greater for the at least bachelor's degree group compared to the bachelor's degree group,  $F(1, 4320) = 10.55$ ,  $p = .001$ ,  $n_p^2 = .002$ , 90% CI [.001, .006]. There was no significant difference between the sexes,  $F(3, 4320) = 0.39$ ,  $p = .531$ ,  $n_p^2 < .001$ , 90% CI [.000, .001], and no interaction between age and sex,  $F(3, 4320) = 0.35$ ,  $p = .791$ ,  $n_p^2 < .001$ , 90% CI [.000, .001].

Pairwise comparisons of ESS scores by age were made using Bonferroni corrections. The only significant differences ( $p < .05$ ) were between 18-25-year olds and 31-35-year-olds, between 26-30-year olds and 31-35-year olds, and between 18-25 and 36-40-year olds.

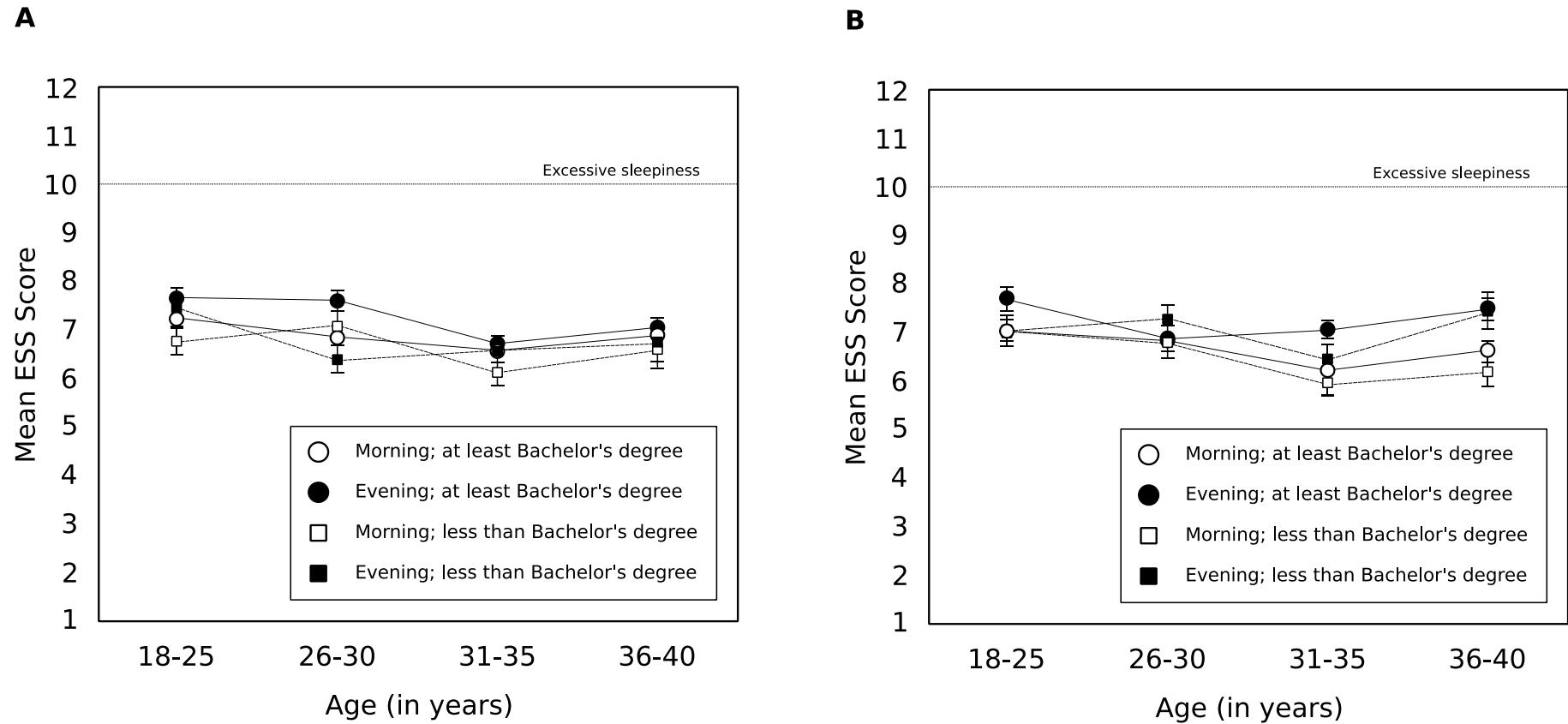


Figure 5. A) Average ESS scores for females by time of day and education. B) Average ESS scores for males by time of day and education.

Error bars represent standard error. Scores above the dotted line would indicate excessive sleepiness.

**Table 7.** ESS scores split by time of day, education, gender, and age category.


	Morning																													
	Less than college educated														At least College Educated															
	Male					Female					Total				Male					Female					Total					
	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3
Age (Yrs)																														
18-25	7.02	3.29	7.00	5.00	9.00	6.79	2.83	7.00	4.00	9.00	6.88	3.05	7.00	5.00	9.00	7.10	3.08	7.00	5.00	9.00	7.33	3.14	8.00	5.00	9.00	7.24	3.11	7.00	5.00	9.00
26-30	6.80	3.20	6.00	5.00	9.75	7.12	3.65	7.00	4.00	10.00	7.01	3.50	7.00	4.00	10.00	6.84	3.53	6.00	4.00	9.00	6.88	3.17	7.00	5.00	9.00	6.86	3.33	6.00	4.00	9.00
31-35	5.96	3.18	6.00	4.00	8.00	6.11	2.92	6.00	4.00	8.00	6.06	3.01	6.00	4.00	8.00	6.22	3.11	5.00	4.00	9.00	6.63	3.15	6.00	4.00	8.00	6.46	3.14	6.00	4.00	8.75
36-40	6.22	2.97	6.00	4.00	8.00	6.59	3.35	7.00	4.00	9.00	6.46	3.21	6.00	4.00	8.00	6.61	3.00	6.00	5.00	9.00	6.89	3.21	6.00	5.00	9.00	6.80	3.14	6.00	5.00	9.00
Total	6.52	3.19	6.00	4.00	9.00	6.62	3.21	7.00	5.00	9.00	6.58	3.20	6.00	4.00	9.00	6.65	3.24	6.00	4.00	9.00	6.88	3.17	7.00	5.00	9.00	6.79	3.20	6.00	5.00	9.00
	Evening																													
	Less than college educated														At least College Educated															
	Male					Female					Total				Male					Female					Total					
	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3
Age (yrs)																														
18-25	7.04	3.12	7.00	5.00	9.00	7.46	3.11	7.00	5.00	10.00	7.24	3.11	7.00	5.00	9.00	7.70	2.99	8.00	6.00	10.00	7.67	3.07	8.00	5.00	10.00	7.68	3.04	8.00	6.00	10.00
26-30	7.29	3.67	7.00	5.00	9.50	6.39	2.84	6.00	4.00	9.00	6.72	3.18	6.00	4.50	9.00	6.91	3.19	7.00	5.00	9.00	7.64	2.96	7.00	5.00	9.00	7.35	3.08	7.00	5.00	10.00
31-35	6.46	3.44	6.00	4.00	9.00	6.56	3.24	7.00	4.00	9.00	6.53	3.29	6.00	4.00	9.00	7.07	3.54	7.00	4.00	9.00	6.71	2.96	7.00	5.00	9.00	6.86	3.22	7.00	4.00	9.00
36-40	7.44	3.06	7.00	6.00	9.00	6.74	3.34	7.00	4.00	9.00	7.00	3.24	7.00	5.00	9.00	7.50	3.17	7.00	6.00	10.00	7.04	3.10	7.00	5.00	9.00	7.18	3.13	7.00	5.00	9.00
Total	7.05	3.28	7.00	5.00	9.00	6.85	3.14	7.00	4.00	9.00	6.93	3.19	7.00	5.00	9.00	7.23	3.27	7.00	5.00	9.50	7.24	3.04	7.00	5.00	9.00	7.24	3.13	6.00	5.00	9.00

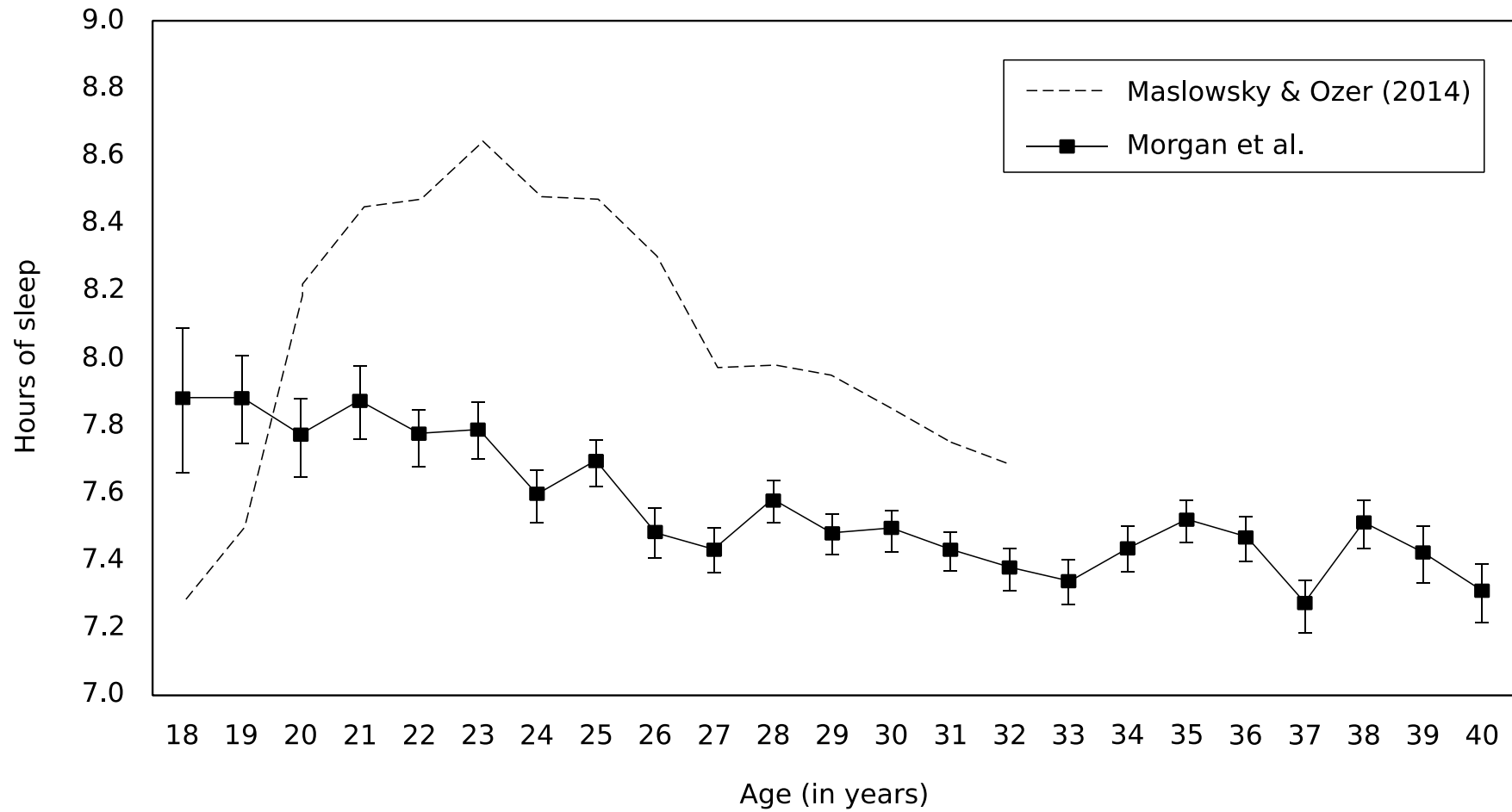
*Sleep Duration: Effects of Age, Time of Day, Education, and Sex*

Figure 6 shows the mean sleep duration as a function of age. The average number of hours slept over the previous night was 7.51 (SD = 1.34). Table 8 shows the breakdown of sleep duration across age group, education, and sex. There was a significant decrease in sleep duration with age,  $F(3, 7468) = 11.48, p < .001, \eta_p^2 = .005, 90\% \text{ CI } [.002, .007]$ . There was no significant effect of on education,  $F(1, 7468) = 1.51, p = .219, \eta_p^2 < .001, 90\% \text{ CI } [.000, .001]$ , sex,  $F(1, 7468) = 0.08, p = .769, \eta_p^2 = .000, 90\% \text{ CI } [.000, .000]$ , or the interaction between sex and age,  $F(1, 7468) = 1.18, p = .314, \eta_p^2 < .001, 90\% \text{ CI } [.000, .001]$  on sleep duration.

Pairwise comparisons of sleep duration by age were made using Bonferroni corrections. There was a significant difference in sleep duration between 18-25-year olds and each of the other age groups,  $p < .001$ . There were no significant differences between the other age groups ( $p > .050$ ).

## Discussion

Normative data of the SSS, ESS, and sleep duration are provided for use in research and clinical settings. Sleep duration was negatively correlated with sleepiness scores on the SSS and the ESS. There were no significant sex differences in sleepiness or sleep duration. There were significant differences in age, time of day, and education in sleepiness and sleep duration. However, all of the significant differences need to be interpreted with caution because the sample sizes are large but most of the effect sizes were trivially small (Sullivan & Feinn, 2012). The only finding with a small to medium effect size was the difference in SSS scores when provided in the morning compared to in the evening.



*Figure 6.* Hours of sleep by age (collapsed across sex and education) in the current study (Morgan et al.) and Maslowsky and Ozer (2014). Error bars represent standard error.

**Table 8.** Sleep duration scores split by time of day, education, gender, and age category.

	Less than college educated															At least College Educated														
	Male					Female					Total					Male					Female					Total				
	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3	M	SD	$\tilde{x}$	Q1	Q3
	Age (Yrs)																													
18-25	7.59	1.43	7.50	7.00	8.50	7.81	1.54	8.00	7.00	9.00	7.70	1.49	8.00	7.00	9.00	7.71	1.27	7.50	7.00	8.50	7.76	1.29	8.00	7.00	8.50	7.74	1.28	8.00	7.00	8.50
26-30	7.51	1.34	7.50	7.00	8.50	7.57	1.45	7.50	6.87	8.50	7.55	1.41	7.50	7.00	8.50	7.46	1.17	7.50	7.00	8.50	7.60	1.18	7.50	7.00	8.50	7.54	1.18	7.50	7.00	8.42
31-35	7.17	1.44	7.50	7.00	8.00	7.48	1.38	7.50	6.50	8.50	7.38	1.41	7.50	6.50	8.17	7.36	1.04	7.50	6.75	8.00	7.51	1.20	7.50	7.00	8.50	7.45	1.14	7.50	7.00	8.00
36-40	7.32	1.20	7.30	6.50	8.00	7.27	1.32	7.50	6.50	8.00	7.29	1.27	7.50	6.50	8.00	7.27	1.11	7.08	6.50	8.00	7.52	1.56	7.50	7.00	8.30	7.44	1.45	7.50	6.75	8.00
Total	7.44	1.39	7.50	6.50	8.17	7.56	1.44	7.50	6.75	8.50	7.51	1.42	7.50	6.50	8.50	7.45	1.15	7.50	7.00	8.00	7.58	1.21	7.50	7.00	8.50	7.53	1.18	7.50	7.00	8.25



### *Time of Day: SSS and ESS*

Participants reported being significantly sleepier if asked in the evening compared to in the morning on the SSS. Given the small-to-medium effect size this difference should be taken into consideration when investigating issues related to sleep. That is, different levels of sleepiness could confound many findings in investigations of impact that sleep has on cognition.

ESS scores were also higher in the evening compared to the morning, though the effect size was very small (unlike differences in SSS scores). Given that the ESS is intended to be a scale which measures sleepiness in “recent times,” the ESS should be unaffected by time of day. This effect may reflect a memory bias when reporting sleepiness (i.e., current sleepiness affects how one remembers general sleepiness) and may need to be considered by researchers and clinicians.

### *Age: SSS, ESS and Sleep Duration*

Sleepiness, as measured by the SSS and ESS, and sleep duration decreased significantly with age but the effects were slight. Previous reports found no differences in daytime sleepiness, as measured by the ESS, across age (e.g., Beaudreau et al., 2012; Gander et al., 2005; Ng & Tan, 2005; Spira et al., 2011; Wu et al., 2012; Sauter et al., 2007). This may reflect differences in samples, including samples with wider age ranges, smaller sample sizes, and the large samples with many more older adults.

In another report in which data were collected from a large sample of adolescents and young adults on sleep duration, the hours slept per night decreased with age (e.g., Maslowsky & Ozer, 2014). In the samples from both studies, we can compare sleep duration from young adults aged 18-32 years (the years of overlap between the two studies). Figure 3 shows the

differences between average hours slept by age for theirs and the current report. In our sample, sleep duration decreased as age increased whereas in the other sample, sleep duration peaked in the early 20s. Our sample also had much less variability over the ages. On average, collapsed across age, our participants reported sleeping 45 minutes less than the other report. The differences between the two studies may account for the different findings. In the Maslowsky and Ozer study, across the years of data collection, four waves of survey questions were administered, and the sleep duration questions changed from “How many hours of sleep do you usually get?” to asking about wake times and bedtimes on weekends and weekdays. To answer these questions, participants had to rely on memory and average their sleep duration across time. Our participants had to rely less on memory and averaging because they were asked, “How long did you sleep last night?” Moreover, the dates of data collection in Maslowsky and Ozer (2014) ranged from 1994-2008 whereas the dates of our data collection ranged from 2017-2019. After smartphone use became popular and two years after Netflix introduced its streaming service, research has demonstrated that usage of technology is greater in the evening (Ahn, Wijaya & Esmero, 2014) and that technology use is associated with disturbed sleep in young adults (e.g., Levenson, Shensa, Sidani, Colditz & Primack, 2016).

*Sex: SSS, ESS and Sleep Duration*

Even with our large sample size where differences can easily be significant (Sullivan & Feinn, 2012), there were no sex differences in SSS and ESS scores or sleep duration. This finding is in line with some previous reports of ESS scores and sex (Ng & Tan, 2005; Wu et al., 2012). Other reports, however, have reported that males have greater daytime sleepiness than females, as measured by the ESS (Bloch, Schoch, Zhang & Russi, 1999; Gander et al., 2005; Joo et al., 2009). Moreover, our finding that there are no sex differences in sleep duration differs from a previous finding that women sleep more in younger adulthood than males

(Maslowsky & Ozer, 2014). While these discrepancies warrant further investigations, we provide normative data broken down by sex.

#### *Education: SSS, ESS and Sleep Duration*

To assess any sleepiness and sleep duration differences across levels of education, we compared the participants with less than a Bachelor's degree to participants with at least bachelor's degree. Sleep duration did not differ between those groups, nor did it differ between males and females, but it did decrease with age. Participants with at least a Bachelor's degree had higher sleepiness scores on both the SSS and ESS than those with less than a Bachelor's degree (although the effect size was very small).

There is a previous report that Americans with lower education (i.e., less than a high school education) had more sleep complaints (Gander et al., 2005). Whether this would translate to higher SSS and ESS scores is unknown because we did not have many participants in our sample with only some high school. In fact, 44% of participants reported having earned a Bachelor's degree, which is highly educated relative to the general population. According to the United States Census Bureau (2019), 22% reported having earned a Bachelor's degree. Because individuals with lower educational attainment are more prevalent in the US and are underrepresented in our sample, further sampling is required.

#### *Limitations*

Although valuable, our large-scale examination of normative values for the SSS, ESS, and sleep duration is not fully representative of the population. Instead it represents young American adults who are highly educated. Additionally, 70% of our sample is Caucasian and therefore cannot assess sleepiness and sleep duration in participants of other ethnicities. We did find time of day differences on SSS and ESS scores, with participants reporting higher levels of sleepiness in the evening. Because we only collected data between 8am-11am and

8pm-11pm, we cannot assess in detail how levels of sleepiness may increase during rest of the day.

### *Conclusion*

We provide normative sleepiness and sleep duration data from a large sample of young American adults. There were differences of age and education, but not sex, on current and daytime reports of sleepiness, and differences of age in sleep duration. All of the effects were trivially small (see Sullivan & Feinn, 2012) except for one. SSS scores provided in the evening were higher – indicating higher sleepiness – than SSS scores provided in the morning. This finding may have implications for researchers investigating the impact of sleep on cognition. The normative SSS and ESS scores and sleep duration of young adults can be used by researchers and clinicians alike.

## Chapter 4: No Effect of Sleep on the Misinformation Effect

### Abstract

Misinformation is increasingly disseminated amongst the public raising great concern over the accuracy of the information people use to guide their decision making in various aspects of their lives (e.g., vaccinations, beliefs about global warming). Sleep is often reported to be beneficial to memory but some investigations in the field of eyewitness memory have found that sleeping after witnessing an event increases the likelihood of accepting new misinformation of the event as being true. It is also possible that this misinformation effect is greater when misinformation is presented before sleep owing to sleep's role in integrating newly encoded memories with each other. We attempted to conceptually replicate and extend previous research by examining sleep's role in integrating true and false information. Participants ( $N = 300$ ), assigned to sleep, wake, or time-of-day control conditions, witnessed a mock crime and were shown misinformation before or after a period of sleep or wake. Participants were subsequently tested on their memory for the original event, misinformation, and source memory. We found that while sleep strengthens memory for the original event, it did not strengthen memory for misinformation or promote its integration with the original event. The data collected in this experiment should be treated with caution since no differences in memory were found between experimental and control conditions. Therefore, we discuss reasons why new methods of measuring the misinformation effect are now needed.

## Introduction

The dissemination of misinformation has increased in recent years and the negative consequences of endorsing this information are of great concern (Ruths, 2019). This phenomenon has affected health care (reduction in vaccination rates has led to increases in measles; Majumder et al., 2015; Public Health England, 2018), perception of climate change (rise in climate change denial and US withdrawal from the Paris Climate Deal; Dunlap, McCright & Yarosh, 2016), the rise of flat earthers (YouGov, 2018), politics (increase in social media bots influencing key political votes such as the EU referendum in 2017; Bastos & Mercea, 2018), and other areas of debate in Western societies. The term “misinformation effect” describes the phenomenon where individuals unwittingly endorse misinformation that was provided to them (Loftus, 1975).

Much of the research on memory for misinformation has been conducted in the domain of eyewitness memory (e.g., Loftus, 1975; Loftus, 2005; Stark, Okado & Loftus, 2010). In these experiments, participants watch a mock crime, are presented with misinformation, and are then tested on their memory for the crime. The misinformation effect occurs when participants endorse the misinformation as having happened during the commission of the crime. For example, during an investigation an investigator may ask the witness if the perpetrator was wearing a red hat. The eyewitness may later incorporate a red hat into their memory as having been present during the commission of the crime, even though it was introduced by the investigator instead. In this example the eyewitnesses would be misattributing information to the wrong source. This type of error is a source monitoring error (Mitchell & Johnson, 2000). It is easy to appreciate how this type of error can jeopardise the integrity of eyewitness evidence. Finding ways to mitigate the susceptibility to source monitoring errors is therefore an important aim. Sleep may be one such source of mitigation.

Sleep is broadly regarded as beneficial for declarative memory (e.g., Diekelmann & Born, 2010; Feld & Born, 2017; Stickgold, 2005). Learning new information followed by a period of sleep is associated with better memory for that information compared to an equivalent period of wake (Drosopoulos, Wagner & Born, 2005; Jones, MacKay, Mantua, Schultz & Spencer, 2018; Maurer et al., 2015; Wilson, Baran, Pace-Schott, Ivry & Spencer, 2012; but see Diekelmann, Wilhelm, & Born, 2009; but see Morgan et al., 2019). Three major processes have been implicated in sleep's benefit on memory. First, sleep may facilitate memory consolidation (i.e., newly learned information is strengthened and stabilized). Second, sleep may enable more efficient encoding (i.e., inputting new information into memory) (Diekelmann & Born, 2010; Feld & Diekelmann, 2015; Mander, Santhanam, Saletin & Walker, 2011; Stickgold, 2005). Third, sleep may integrate new information with existing information (Tamminen et al., 2010; Tamminen et al., 2013). These reasons have been proposed as explanation for the better memory performance observed after sleep.

In standard memory and sleep experiments, the information presented and tested is typically lists of images or words and conflicting information is not introduced and tested (e.g., Drosopoulos, Wagner & Born, 2005; Jones, MacKay, Mantua, Schultz & Spencer, 2018). If integrating information is aided by sleep processes, then it is possible that conflicting information would be integrated and therefore lead to lower memory performance. If so, then integration is most likely to occur if the information and misinformation is presented, and then a period of sleep takes place. And if sleep does help increase memory performance by enhancing consolidation and/or encoding, and if a period of sleep takes place between the original information and misinformation, the original memory should be protected from the conflicting information resulting in greater memory performance than if sleep had not occurred.

There are two sleep experiments in which conflicting information was introduced. In one experiment, participants were given the Gudjonsson Suggestibility Scale (van Rijn, Carter,

McMurtrie, Wilner, & Blagrove, 2017). The Gudjonsson Suggestibility Scale measures susceptibility to misinformation and works as follows: participants read a story, freely recall the story, then after an interval, they are presented with misinformation in the form of misleading questions, and then freely recall the story again (Gudjonsson, 1984). In van Rijn et al., participants were assigned to a sleep or wake condition and therefore slept or remained awake during the interval. Overall memory was better in the sleep than wake condition but there was no difference in endorsing the misinformation as having been part of the original story.

In another experiment in which the goal was to investigate the impact of sleep on memory for misinformation, participants were shown a series of photos depicting the events of a mock crime, they received information that differed from the information in the photos of the crime (referred to as misinformation), and took a recognition memory test (Calvillo, Parong, Peralta, Ocampo, & Van Gundy, 2016). There were four conditions that varied by the duration between the presentation of the mock crime photos and receiving the misinformation. The duration was of 12 hours of daytime wake (i.e., did not include a night of sleep), 12 hours overnight (i.e., included one night of sleep), 24 hours (i.e., included one night of sleep), or no duration (i.e., the misinformation was presented immediately after the presentation of the mock crime photos). Thus, the conditions were 12-hour wake, 12-hour sleep, 24-hour sleep, and immediate, respectively. During the recognition memory test, participants were tested on target items (items that were in the original photos), lure items (items that were neither in the original photos or part of the misinformation stimuli, and misinformation items (items that were presented after viewing the mock crime photos).

As shown in Figure 1 of Calvillo et al. (2016), participants in the sleep condition were better able to discriminate targets from lures in the immediate condition than the other conditions (i.e., better old/new recognition memory). Participants in both sleep conditions were



better able to discriminate targets from lures than participants in the wake conditions. There were no differences in discriminability in the sleep conditions (i.e. 12 hour sleep retention interval vs 24 hour retention interval with sleep). Participants in the sleep conditions endorsed misinformation items more than participants in the wake and immediate conditions<sup>8</sup>.

In accordance with predictions based on the sleep literature, the better old/new recognition memory for participants in the sleep conditions indicates that they consolidated the original information. However, based on the sleep literature, they should have encoded the misinformation better and thus been able to reject the misinformation as having been part of the original mock crime scenario, but they did not. Because the misinformation was presented after sleep, sleep could not have played a role in integrating misinformation and original information. That is, the misinformation should have been rejected more in the sleep conditions as there was not a period of sleep to consolidate and integrate the misinformation with the originally presented information and because sleep should enable better encoding.

The fact that participants did not reject the misinformation as having been part of the original episode may mean that they remembered having been presented with both types of information (which they were) but were less able to remember the sources of the information. Including a source memory test would validate this speculation. Moreover, whether a period of sleep after the misinformation was presented would result in more source confusion because whether the different types of information would be integrated is unknown. Based on prior literature, it is a reasonable hypothesis.

For instance a related phenomenon is that if people recall information inaccurately, they later remember their less accurate memory over the correct information. This is referred to as Retrieval Induced Distortion (RID; Bridge & Paller, 2012). Cairney, Lindsay, Paller and

---

<sup>8</sup> This is based on d' scores computed from misinformation and lure values derived from Calvillo et al. (2016) Figure 1.

Gaskell (2018) examined whether sleep would increase the likelihood of RID occurring, as predicted based on the role that sleep has on consolidation. In that experiment participants first passively viewed a series of words appearing on specific locations in a grid. Next participants actively learnt the locations of the words by attempting to drag the words to the locations they were originally presented in. This continued until participants reached an accuracy criterion. Then participants completed the first testing session on all of the word locations that they had just learnt which was followed by an interval of sleep or wake. Lastly participants completed a second testing session again on all of the word location pairs. After a period of sleep, memory for the word locations at the second testing session were closer to those made during the first testing session and the originally studied locations compared to a period of wake.

These results were taken as support for the Multiple Trace Transformation model of memory, which holds that when information is encoded, a single memory trace for that information exists (MTT; Moscovitch & Nadel, 1998; Nadel & Moscovitch, 1997; Winocur et al., 2010). If this information is retrieved, then a second memory trace of that same information will be formed, in this case one that is not erroneous but more similar to what was originally studied. Cairney et al. (2018) proposed that the RID produced a second memory trace at the first testing session that was subsequently consolidated with the original memory trace during sleep. In addition to consolidating the two memory traces, Cairney et al. (2018) suggested sleep may have also integrated them. This may happen when misinformation is presented. In other words the misinformation may be equivalent to the memory trace of the first retrieval session in the Cairney et al. (2018). Once it has undergone consolidation, subsequent retrieval of the misinformation after may be greater because sleep preserved the memory trace for original information and the misinformation just as it preserved the memory trace for the studied locations and retrieved locations in Cairney et al. (2018).

Whether sleep mitigates or harms memory for misinformation would largely depend on the time at which the misinformation is presented within the wake/sleep cycle. To test this, we adapted the procedure used by Calvillo et al. (2016). We added a source memory test and the conditions were different. There was no 24-hr sleep condition, we added a 12-hr wake and 12-hr sleep condition after misinformation was presented, and we added two time-of-day control conditions. All participants witnessed two events, then received misinformation about them and were given recognition memory tests and a source memory test. A retention interval (sleep vs. wake) was included either before or after receiving the misinformation. Thus there were four experimental conditions, see Figure 7: sleep before misinformation, wake before misinformation, sleep after misinformation and wake after misinformation. There were three research questions regarding how the placement of misinformation before or after sleep affects old/new recognition memory, memory for misinformation, and source memory. The research questions, hypotheses, method, and data analysis plans were pre-registered and corresponding data are available at <https://osf.io/e85xy/>. Below we outline the hypothesis for each research question.

#### *Old/new recognition memory*

Our first research question was: Does the presentation of misinformation before or after a retention interval of sleep or wakefulness impact old/new recognition memory performance? Based on the wider sleep literature, we predicted that if participants slept after receiving misinformation, they would have better old/new recognition memory compared to participants who remained awake after receiving misinformation. Those who slept after receiving the misinformation would consolidate the original information (the old/new items) more so than those who remained awake. For the same reason and based on our reanalysis of Calvillo's et al. (2016) findings, we also predicted that if participants slept before receiving misinformation,

they would have greater old/new recognition memory compared to those who remained awake before receiving misinformation.

### *Memory for misinformation*

Our second research question was: Does the presentation of misinformation before or after a retention interval of sleep or wakefulness impact memory for misinformation? Based on the wider sleep literature, we predicted that participants who sleep after receiving misinformation will endorse misinformation more compared to participants who remain awake after receiving misinformation. Those who slept after receiving misinformation may integrate the original and misinformation during sleep. For this reason, we also predicted that participants who sleep after receiving misinformation will endorse the misinformation more compared to those who sleep before receiving misinformation. Moreover, based on the findings reported in Calvillo et al. (2016), we predicted that participants who sleep before receiving misinformation would be more likely to endorse misinformation compared to participants who remained awake before receiving misinformation.

### *Source memory*

Our third research question was: Does the presentation of misinformation before or after a retention interval of sleep or wakefulness impact source memory? We predicted that source memory will be greater for participants who remained awake after receiving misinformation compared to those who slept after receiving misinformation. This prediction is based on the RID findings that sleep may consolidate two memory traces simultaneously and integrate the erroneous traces for the misinformation into their memory for the original event leading to poorer source memory performance (Cairney et al., 2018). For the same reason, we predicted that the sleep before misinformation condition will have greater source memory compared to the sleep after misinformation condition. Based on the findings reported in Calvillo et al.

(2016), that the participants who slept before receiving misinformation endorsed more of it, we predicted that source memory will be greater for participants who remained awake before receiving misinformation compared to participants who slept before receiving misinformation. By testing these hypotheses, we can assess the potential role and timing of sleep in strengthening memory for both inaccurate and accurate information, and whether sleep integrates both types of information into a composite memory trace.

## Methods

### *Participants*

Sensitivity analysis indicated that that by setting the parameters to standard values of  $\alpha = 0.05$  and  $1 - \beta = 0.80$ , 50 participants were needed in each condition to detect an effect size of  $f = 0.24$ . Three-hundred participants were randomly assigned to a condition. Participants who signed up to take part in the morning were randomly assigned to a wake condition or the AM time-of-day control condition and participants who signed up to take part in the evening were randomly assigned to a sleep condition or PM time-of-day control condition. The four experimental conditions were the Sleep After Misinformation ( $n = 50$ ), Wake After Misinformation ( $n = 50$ ), Sleep Before Misinformation ( $n = 50$ ), and Wake Before Misinformation ( $n = 50$ ) conditions. The two time-of-day control conditions were the AM Control ( $n = 50$ ) and PM Control ( $n = 50$ ) conditions. To be eligible to be included in the analyses participants had to answer no to the following questions: Are you currently diagnosed with any sleep disorder (e.g. insomnia, sleep apnoea, etc.)? Are you currently diagnosed with a psychiatric disorder (e.g. depression, posttraumatic stress disorder, etc.)? Are you currently diagnosed with a neurological disorder (e.g. mild cognitive impairment, Alzheimer's disease, etc.) which affects memory? Do you work as a shift worker? Have you travelled across time zones within two weeks? Are you currently taking any prescribed medication that may affect

sleep or memory? Participants were undergraduate students recruited from the University of California, San Diego (UCSD) Department of Psychology. Demographic information is presented in Table 1. The UCSD Internal Review Board approved this study (#172091).

### *Materials*

*Three-phase misinformation paradigm.* We adapted a version (Calvillo et al., 2016) of the three-phase misinformation paradigm (Okado & Stark, 2005). The paradigm consisted of a study phase, misinformation phase, and test phase. The study phase included two sets of 50 images depicting a mock crime (one computer theft and the other a purse theft). Each image was presented for 3500 ms and was followed by a fixation cross presented for 500 ms. The misinformation phase included two narratives of the two events they saw in the study phase. For both narratives, 50 sentences were presented for 4000 ms, each followed by a fixation cross presented for 500 ms. Sentences longer than 4 lines were presented for 4500 ms to make sure participants had time to read them. Twenty-four sentences in each narrative contained information that was inconsistent with what they were shown in the study phase. For example, if in the images the musician character entered a rehearsal room and expressed surprise at the lack of a piano, the inconsistent misinformation sentence might read “The musician was surprised to see no piano stool”, while another participant would see consistent information in the form of “The musician was surprised to see no piano”. The remaining sentences contained information that was consistent with the images.

Test phase items were 72 statements about the two events that participants were shown in the study phase. These 72 statements included 24 target items (i.e., old items), 24 misinformation items (i.e., information presented *only* during the misinformation phase) and 24 lure items (i.e., new items not presented in either the study or misinformation phases). There were 36 statements for each of the events. The statements within each event were presented in random order. For target, misinformation, and lure items 12 items corresponded to the first

mock crime and 12 correspond to the second mock crime. For each statement it was required to indicate whether a statement did or did not occur in the pictures viewed in the study phase. Responses were a “yes” if the statement did occur in the study phase, and a “no” if it did not. Confidence ratings were collected on a 3-point scale (e.g., 1 = Guess, 2 = Sure, 3 = Very sure), although these ratings are not analyzed here.

*Source memory task.* Items for the source memory task were the 48 statements that participants were shown in the recognition memory task. These statements consisted of the 24 target items and the 24 misinformation items. For both target and misinformation items 12 corresponded to the first mock crime and 12 corresponded to the second mock crime. Participants were asked to indicate whether they thought the statement occurred in the pictures shown in the study phase or the sentences shown in the misinformation phase on a 6-point scale. The rating options were “1 (definitely in the sentences)”, “2 (probably in the sentences)” or “3 (maybe in the sentences)” if the participant thought the statement occurred in the misinformation phase. For pictures responses were “4 (maybe in the pictures)”, “5 (probably in the pictures)” or “6 (definitely in the sentences)” if the participant thought the statement occurred in the study phase.

*Distractor task.* The distractor task was 25 anagrams of US states. This task was relevant to participants in the control conditions in order to introduce a short gap between the study phase and the misinformation phase and the misinformation phase and recognition and source memory tasks. This task was included in all conditions for the sake of consistency across experimental and control conditions.

*Sleep related questions.* The questions were, when did you wake up, what time did you go to sleep and how long did you sleep for?

*Stanford Sleepiness Scale (SSS)*. The SSS measures the current level of sleepiness (Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973). The SSS is 7-point scale ranging from 1 = feeling active, vital, alert or wide awake to 7 = no longer fighting sleep, sleep onset soon; having dream like thoughts. High scores indicate high levels of sleepiness and low scores indicate low levels of sleepiness.

*Reduced Morningness Eveningness Questionnaire (rMEQ)*. The rMEQ (Adan & Almirall, 1991) measures chronotype (i.e., more alert in the morning or evening) with five questions, which are: What time would you get up if you were free to plan your day? During the first half-hour after you wake up in the morning, how tired do you feel? At what time of the day do you feel you become tired as a result of need for sleep? At what time of day do you think that you reach your "feeling best" peak? One hears about "morning" and "evening" types of people. Which one of these types do you consider yourself to be? Low scores indicate evening types and high scores indicate morning types. Results from this questionnaire are not analysed here.

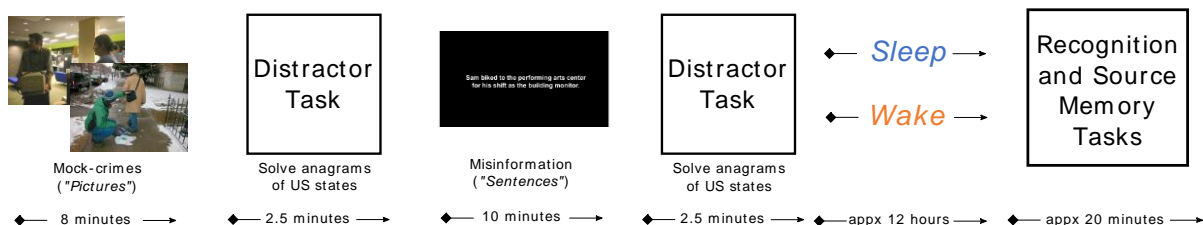
*Procedure*. We used the AM-PM design commonly used by sleep researchers (e.g., Drosopoulos, Wagner, & Born, 2005; Payne et al., 2008). In the AM-PM design participants are assigned to a wake (AM-PM) or sleep condition (PM-AM). In our integrated design, participants in the sleep conditions completed the study phase in the evening (PM, between 8pm-11pm) and the test phase in the morning (AM, between 8am-11am), while those in the wake conditions completed the study phase in the morning (AM, between 8am-11am) and test phase in the evening (PM, between 8pm-11pm). Figure 7 shows the procedure for the sleep after misinformation, wake after misinformation, sleep before misinformation and wake before misinformation conditions. Timing of the misinformation phase depended on the condition. In the sleep after misinformation condition the misinformation phase took place in the evening immediately after the study phase and the distractor task. In the wake after misinformation



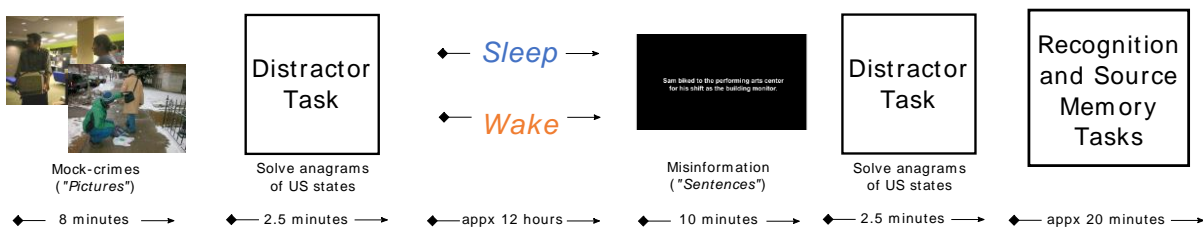
condition the misinformation phase took place in the morning immediately after the study phase and the distractor task. In the sleep before misinformation condition the misinformation phase took place in the morning after the evening study phase. In the wake before misinformation condition the misinformation phase took place in the evening after the morning study phase. For time-of-day controls the study phase, misinformation phase and test phase occurred in one continuous session in the morning (AM, between 8am-11am) or in the evening (PM, between 8pm-11pm).

Participation took place online. All participants first completed the consent form. Then participants answered the exclusion questions, sleep-related questions, and completed the SSS. Participants then began the three-phase misinformation paradigm described above. During the study phase and the misinformation phase participants were shown the pictures/statements for two mock crimes: a computer theft, and then a purse theft. In between seeing the pictures and

#### Sleep/wake after misinformation



#### Sleep/wake before misinformation



AM/PM

Figure 7. Procedure for the experimental conditions.

statements and after seeing the statements participants completed the distractor task for 2.5 minutes. After the study phase participants in the wake condition were instructed not to nap between the study and test phases, and were asked at the beginning of the test phase whether they did nap. During the test phase all participants saw the statements for the computer theft and then another video of a purse theft. Next, participants completed the source memory task and then the SSS and the rMEQ. After completing the sleep questionnaires and sleep-related questions participants answered a validation question about the video (“What were the crimes committed in the two events?”) to ensure that attention was paid during the study phase. Participants were also asked “Have you seen the images presented to you in this experiment prior to participating” to ensure none of the participants had taken part in previous experiments using the same stimuli. Finally, participants were debriefed.

## Analysis Strategy

### *Confirmatory analyses*

Old/new memory, memory for misinformation, and source memory were analysed with a one-way ANOVA with experimental condition (i.e., Sleep After Misinformation, Wake After Misinformation, Sleep Before Misinformation, Wake After Misinformation) as the independent variable and  $d_a$  as the dependent variable.  $d_a$  is a measure of discriminability to determine the ability to discriminate between events and items that did or did not occur in the study phase and the sources of those events (i.e., the study phase or misinformation phase).  $d_a$  is given by

$$d_a = \frac{(\mu_t - \mu_l)}{\sqrt{\frac{\sigma_t^2 + \sigma_l^2}{2}}}$$

where  $\mu_t$  refers to the proportion of “yes” responses made to targets (i.e. a hit) and  $\mu_l$  refers to the proportion of “yes” responses made to lures (i.e. a false alarm).  $\sigma_t^2$  and  $\sigma_l^2$  refer to the standard deviations of the proportion of “yes” responses made to targets and lures.  $d_a$  is similar to the more commonly used  $d'$ , but accounts for the possibility that the standard deviations of the distributions may not be equal (MacMillan & Creelman, 2005).

For old/new recognition memory,  $d_a$  determined the ability to distinguish targets from lures. “Yes” responses to statements that were consistent with the witnessed events were counted as hits, and “yes” responses to lures were counted as false alarms. Higher  $d_a$  therefore indicates more accurate old/new recognition memory. For memory for misinformation,  $d_a$  determined the ability to distinguish misinformation from lures. “Yes” responses to statements that were consistent with the misinformation were counted as hits, and “yes” responses to lures were counted as false alarms. Note that because “yes” responses to misinformation were counted as hits, in this category of memory higher  $d_a$  indicates higher propensity to endorse misinformation as true information.

For source memory  $d_a$  determined the ability to distinguish information that occurred in the study phase from information that occurred in the misinformation phase. To this end we recoded responses made with options 1 to 3 as “picture” responses, and responses made with options 4 to 6 as “sentences” responses. “Picture” responses made to statements that were consistent with the true information were counted as hits, while “picture” responses to statements that only occurred in the misinformation phase were counted as false alarms. Items which represented  $\mu_t$ ,  $\mu_l$ ,  $\sigma_t^2$  and  $\sigma_l^2$  varied depending on the type of memory being examined, see Appendix A. Orthogonal planned contrasts were used to test each hypothesis and compare  $d_a$  between conditions for all research questions. Alpha levels for each hypothesis test were set to  $\alpha = 0.05$ .

*Secondary analyses*

To rule out time-of-day effects we compared  $d_a$  between the two time-of-day control conditions (AM vs. PM). Three independent measures  $t$ -tests were used to compare AM and PM  $d_a$  for old/new recognition memory, memory for misinformation, and source memory. Sensitivity analysis for an independent measures  $t$ -test indicated that by setting the parameters to standard values of  $\alpha = 0.05$  and  $1 - \beta = 0.80$ ,  $n = 50$  per condition can detect an effect size of  $d = 0.56$  in a two-tailed test. This sensitivity analysis informed data collection for all secondary analyses.

**Results**

A total of 443 participants took part, but 143 of those participants were excluded as per our exclusion criteria specified in our pre-registration. Participants were excluded from the analysis for reporting that they have a sleep disorder ( $n = 36$ ), they have travelled across time zones in the past two weeks ( $n = 32$ ), they are currently taking medication that may affect their sleep ( $n = 20$ ) or their memory ( $n = 6$ ), they have a psychiatric disorder ( $n = 29$ ) or neurological disorder ( $n = 2$ ), had seen the images presented in this experiment before participating ( $n = 5$ ), incorrectly answered the validation question ( $n = 11$ ), or, for the wake condition, napped between study and test ( $n = 18$ ). Data from the remaining 300 participants were used to carry out the following analyses (see Table 9 for demographic information). Table 10 shows the  $d_a$  scores by experimental condition for old/new recognition memory, memory for misinformation, and source memory.

*Confirmatory analyses**Old/new recognition memory*

A one-way between-subjects ANOVA revealed a significant difference between conditions on  $d_a$ ,  $F(3, 196) = 2.71$ ,  $p = .047$ ,  $f = 0.16$ , 95% CI: [0.00, 0.32]). Planned comparisons revealed that those in the sleep after misinformation condition had significantly higher old/new recognition memory performance compared to those in the wake after misinformation condition ( $p = .025$ ). Other planned comparisons revealed no significant difference in old/new recognition memory performance between sleep before misinformation and wake before misinformation conditions ( $p = .096$ ).

#### *Memory for misinformation*

The one-way between-subjects ANOVA did not reveal a significant difference between conditions on memory for misinformation,  $F(3, 196) = 2.43$ ,  $p = .066$ ,  $f = 0.14$ , 95% CI: [0.00, 0.30]. Planned comparisons yielded no significant differences between sleep after misinformation and wake after misinformation conditions ( $p = .110$ ), or sleep before misinformation and wake before misinformation conditions ( $p = .353$ ).

**Table 9.** Demographic characteristics of participants per condition.

	Condition					
	Sleep before misinformation	Sleep after misinformation	Wake before misinformation	Wake after misinformation	AM Control	PM Control
Gender						
Female	36	30	36	36	36	37
Male	14	18	14	14	14	13
Do not wish to state	0	2	0	4	0	0
Age						
Years ( <i>SD</i> )	20.76 (2.61)	19.84 (1.45)	20.44 (1.45)	20.76 (1.49)	20.86 (2.91)	20.54 (1.56)
Ethnicity						
African American	1	1	0	2	0	1
Arab American	2	1	2	1	1	3
Asian American	22	26	24	27	25	25
Caucasian American	5	4	9	3	7	6
Hispanic American	16	11	11	14	9	11
Indian American	2	2	0	1	5	3
Native American	0	0	0	0	0	0
Do not wish to state	1	1	0	0	0	1
Other	1	4	4	2	3	0
Education						
Bachelor's Degree	10	4	9	7	8	7
High School/GED	10	14	9	11	14	6
Master's Degree	0	0	0	0	0	0
Post-Masters	0	0	0	0	0	0
Some College	29	32	32	31	27	35
Some High School	0	0	0	0	0	0
Do not wish to state	1	0	0	1	1	2

Mean, *M*; Standard deviation, *SD*; Confidence intervals, *CI*

### *Source memory*

A one-way between-subjects ANOVA revealed a significant main effect of condition on source memory,  $F(3, 196) = 2.94$ ,  $p = .034$ ,  $f = 0.17$ , 95% CI: [0.00, 0.33]. Planned comparisons indicated that those in the sleep after misinformation condition had significantly greater source memory than those in the sleep before misinformation condition ( $p = .038$ ). Further comparisons yielded no significant differences between the sleep after misinformation and wake after misinformation conditions ( $p = .433$ ) or the sleep before misinformation and wake before misinformation conditions ( $p = .077$ ).

**Table 10.** Means, standard deviations and 95% confidence intervals for each type of memory per experimental condition.

Type of Memory	<i>M</i>	<i>SD</i>	95% CI
<i>Old/new recognition memory</i>			
Sleep before misinformation	1.04	0.39	[0.91, 1.16]
Sleep after misinformation	1.04	0.43	[0.92, 1.16]
Wake before misinformation	0.84	0.44	[0.71, 0.96]
Wake after misinformation	0.89	0.50	[0.76, 1.01]
<i>Memory for Misinformation</i>			
Sleep before misinformation	1.06	0.57	[0.89, 1.22]
Sleep after misinformation	1.18	0.61	[1.02, 1.35]
Wake before misinformation	0.87	0.42	[0.70, 1.03]
Wake after misinformation	1.07	0.72	[0.91, 1.24]
<i>Source Memory</i>			
Sleep before misinformation	0.72	0.65	[0.55, 0.89]
Sleep after misinformation	0.98	0.65	[0.81, 1.16]
Wake before misinformation	0.63	0.59	[0.45, 0.80]
Wake after misinformation	0.76	0.60	[0.58, 0.94]

### *Time-of-day Controls*

See Table 11 for the  $d_a$  scores for the time-of-day control conditions for old/new recognition memory, memory for misinformation, and source memory. Independent  $t$ -tests revealed no differences in old/new recognition memory,  $t(98) = 0.06$ ,  $p = .956$ ,  $d = 0.02$ , 95%

CI: [-0.37, 0.41], memory for misinformation,  $t(98) = 0.58$ ,  $p = .565$ ,  $d = 0.13$ , 95% CI: [-0.37, 0.41], or source memory,  $t(98) = 0.59$ ,  $p = .553$ ,  $d = 0.11$ , 95% CI: [-0.27, 0.51], between the AM and PM control conditions (Table 11).

**Table 11.** Means, standard deviations and 95% confidence intervals for each type of memory per control condition.

Type of Memory	<i>M</i>	<i>SD</i>	95% CI
<i>Old/new recognition memory</i>			
AM Control	1.06	0.50	[0.92, 1.19]
PM Control	1.05	0.45	[0.92, 1.19]
<i>Memory for Misinformation</i>			
AM Control	1.15	0.65	[0.98, 1.32]
PM Control	1.08	0.51	[0.92, 1.25]
<i>Source Memory</i>			
AM Control	0.81	0.51	[0.65, 0.96]
PM Control	0.87	0.57	[0.72, 1.02]

Mean, *M*; Standard deviation, *SD*; Confidence intervals, CI.

### *Exploratory analyses*

A set of exploratory analyses were carried out to evaluate potential confounding factors based on sleepiness, and to investigate differences in memory between experimental (collapsed across Sleep After Misinformation, Wake After Misinformation, Sleep Before Misinformation, Wake Before Misinformation conditions) and control conditions (collapsed across AM and PM control). We also measured if there is an overall benefit of sleep on old/new recognition memory, memory for misinformation, and source memory when collapsed across conditions. Table 12 shows  $d_a$  scores for all of the memory tests.

### *Old/new recognition memory*

A 2 (sleep vs. wake) x 2 (before misinformation vs. after misinformation) between-subjects ANOVA revealed participants in the sleep conditions had significantly higher old/new recognition memory ( $d_a$ ) than participants in the wake conditions,  $F(1, 196) = 7.76$ ,  $p = .006$ ,  $\eta_p^2 = .038$ , 90% CIs [0.04, 0.16]. There was not a significant difference between participants



who that saw the misinformation before or after sleep or wake,  $F(1, 196) = 0.18, p = .671, \eta_p^2 = .001$ , 90% CIs [0.00, 0.02]. There was also not an interaction between the conditions and the timing of the misinformation,  $F(1, 196) = 0.17, p = .675, \eta_p^2 = .001$ , 90% CIs [0.00, 0.02].

#### *Memory for misinformation*

A 2 (sleep vs. wake) x 2 (before misinformation vs. after misinformation) between-subjects ANOVA revealed no significant differences in memory for misinformation ( $d_a$ ) between participants in the sleep and wake conditions,  $F(1, 196) = 3.22, p = .074, \eta_p^2 = .016$ , 90% CIs [0.00, 0.06], or between participants who that saw the misinformation before or after sleep or wake,  $F(1, 196) = 3.85, p = .051, \eta_p^2 = .019$ , 90% CIs [0.00, 0.06], and no significant interaction between the conditions and the timing of the misinformation,  $F(1, 196) = 0.23, p = .634, \eta_p^2 = .001$ , 90% CIs [0.00, 0.02].

#### *Source memory*

A 2 (sleep vs. wake) x 2 (before misinformation vs. after misinformation) between-subjects ANOVA revealed no significant differences in source memory between the sleep and wake conditions,  $F(1, 196) = 3.28, p = .072, \eta_p^2 = .016$ , 90% CI [0.00, 0.05]. Source memory was better for participants who received the misinformation after the retention interval compared to those who received the misinformation before the retention interval (collapsed across the sleep and wake conditions),  $F(1, 196) = 5.05, p = .026, \eta_p^2 = .025$ , 90% CI [0.00, 0.07]. There was no significant interaction between the conditions and the timing of the misinformation,  $F(1, 196) = 0.49, p = .685, \eta_p^2 = .002$ , 90% CI [0.0, 0.03].

**Table 12.** Means, standard deviations and 95% confidence intervals for each type of memory.

Type of Memory		<i>M</i>	<i>SD</i>	95% CI
<i>Old/new recognition memory</i>				
Sleep vs. Wake				
	Wake	0.86	0.47	[0.77, 0.95]
	Sleep	1.04	0.41	[0.95, 1.13]
Before vs. After				
	Before	0.94	0.43	[0.85, 1.03]
	After	0.96	0.47	[0.87, 1.05]
<i>Memory for Misinformation</i>				
Sleep vs. Wake				
	Wake	0.97	0.59	[0.85, 1.08]
	Sleep	1.12	0.59	[1.00, 1.24]
Before vs. After				
	Before	0.96	0.51	[0.84, 1.08]
	After	1.13	0.66	[1.01, 1.24]
<i>Source Memory</i>				
Sleep vs. Wake				
	Wake	0.69	0.60	[0.57, 0.82]
	Sleep	0.85	0.66	[0.73, 0.98]
Before vs. After				
	Before	0.67	0.62	[0.55, 0.79]
	After	0.87	0.63	[0.75, 0.99]

Mean, *M*; Standard deviation, *SD*; Confidence intervals, CI.

#### *Stanford Sleepiness Scale (SSS)*

The time-of-day control conditions suggest there was no difference between the performance of participants in the morning and in the evening, thus ruling out possible time-of-day confounds in our sleep vs. wake manipulation. We also collected sleepiness ratings from the SSS in the study and test phase, and therefore use these to rule out possible differences in sleepiness between conditions. First, looking at the condition where misinformation was given before the retention interval, there was no significant difference between the sleep group ( $M = 3.16$ ,  $SD = 1.28$ ) and the wake group ( $M = 2.82$ ,  $SD = 1.07$ ) in the study phase,  $t(98) = 1.42$ ,  $p = .158$ ,  $d = 0.28$ , 95% CI: [-0.11, 0.68], or between the sleep ( $M = 3.18$ ,  $SD = 1.34$ ) and wake ( $M = 3.46$ ,  $SD = 1.40$ ) groups in the test phase,  $t(98) = 1.01$ ,  $p = .314$ ,  $d = 0.20$ , 95% CI: [0.13, 0.92]. Turning to the condition where misinformation was given after the retention interval, there was no significant difference between the sleep group ( $M = 2.78$ ,  $SD = 1.10$ ) and the wake

group ( $M = 2.70$ ,  $SD = 0.90$ ) in the study phase,  $t(98) = 0.39$ ,  $p = .695$ ,  $d = 0.08$ , 95% CI: [-0.31, 0.47]. In the test phase there was a significant difference between the sleep ( $M = 3.12$ ,  $SD = 1.31$ ) and wake ( $M = 3.74$ ,  $SD = 1.56$ ) groups,  $t(98) = 2.13$ ,  $p = .035$ ,  $d = 0.43$ , 95% CI: [0.83, 0.03], whereby the wake group was sleepier.

We also compared sleepiness scores in the time-of-day control groups. In the study phase, there was no difference between the AM ( $M = 3.10$ ,  $SD = 1.20$ ) and PM ( $M = 3.18$ ,  $SD = 1.08$ ) groups,  $t(98) = 0.34$ ,  $p = .731$ ,  $d = 0.07$ , 95% CI: [0.37, 0.41]. In the test phase the PM ( $M = 4.18$ ,  $SD = 1.27$ ) group was significantly sleepier than the AM ( $M = 3.42$ ,  $SD = 1.44$ ) group,  $t(98) = 2.76$ ,  $p = .006$ ,  $d = 0.56$ , 95% CI: [0.15, 0.96].

## Discussion

We tested multiple hypotheses for three separate research questions, which were all related to whether the time at which misinformation is presented during the sleep/wake cycle impacts memory (old/new recognition memory, memory for misinformation, and source memory). Based on prior literature sleep could have either mitigating or harmful effects on memory for misinformation, or both. Sleep processes aid consolidation, encoding, and integration of information (Diekelmann & Born, 2010; Stickgold, 2005). Of the eight hypotheses, one was supported: old/new recognition memory was better for participants in the Sleep After Misinformation condition compared to participants in the Wake After Misinformation condition. Our exploratory analyses suggested that this may be a global effect of sleep that is not dependent on the timing of when misinformation is presented.

We included time-of-day controls to ensure that any differences that we may find between the sleep and the wake conditions were not due to a time of day confound. No such confounds were found. We did not however expect to find that our time of day control participants, who had a retention interval of 2.5 minutes, would have similar performance to

the participants in the experimental conditions. An exploratory independent measures *t*-test indicated that there were no significant differences between the control and experimental conditions in old/new recognition memory ( $t(298) = 1.84, p = .066, d = .22$ ); memory for misinformation ( $t(298) = .10, p = .319, d = .12$ ); or source memory ( $t(298) = .87, p = .360, d = .11$ ) (see Table 13). The fact that the scores did not significantly differ across conditions with a 12-hour retention interval and a 2.5-minute retention interval is puzzling and conclusions made from the results of the experimental conditions need to be considered with caution.

**Table 13.** Means, standard deviations and 95% confidence intervals for each type of memory per experimental (collapsed across sleep and wake) and control (collapsed across AM and PM) condition.

Type of Memory	<i>M</i>	<i>SD</i>	95% CI
<i>Old/new recognition memory</i>			
5-minutes	1.06	0.50	[0.92, 1.19]
12-hours	1.05	0.45	[0.92, 1.19]
<i>Memory for Misinformation</i>			
5-minutes	1.15	0.65	[0.98, 1.32]
12-hours	1.08	0.51	[0.92, 1.25]
<i>Source Memory</i>			
5-minutes	0.81	0.51	[0.65, 0.96]
12-hours	0.87	0.57	[0.72, 1.02]

Mean, *M*; Standard deviation, *SD*; Confidence intervals, CI.

This result is likely not due to some idiosyncrasy in our experiment or with our participants. Similarly, in another experiment in which the three-phase misinformation paradigm is used, there was no difference in memory of participants who were sleep deprived and those who were rested (Lo, Chong, Ganesan, Leong & Chee, 2016). Although the retention interval was the same for these conditions in that study, it is reasonable to assume that memory should be worse for the events when participants are tested when they are sleep deprived, compared to when they are rested. These results suggest that there may be something fundamentally problematic about this paradigm.

*Differences between the current findings and the findings of Calvillo et al. (2016).*

In the first investigation of its kind, Calvillo et al. (2016) examined the impact of sleep on the memory for misinformation. Because one of our questions related to integration that happens with sleep, we added another two conditions, removed one (the 24-hours), and added two time-of-day controls. Some of the results differed. In Calvillo et al., participants in the sleep condition (who received misinformation after a period of sleep) were more likely to endorse the misinformation and were poorer at discriminating between misinformation and lures. We, on the other hand, found that no effect of condition on memory for misinformation. Our data do not conceptually replicate Calvillo's et al. findings, and consequently do not support either explanation provided for their misinformation data. They proposed that sleep had extracted the gist of the mock-crime at the expense of the veridical detail (e.g., what the criminal was wearing), so the original memory was more susceptible to distortion through misinformation.

Secondly, they proposed that sleep enables improved encoding of new information, thus leading to a greater memory for misinformation since the misinformation was presented after sleep (Mander et al., 2011). If sleep promoted extraction of gist based representations of the original event, and rendered memory of the original event and later misinformation less distinguishable, we would expect to observe higher  $d_a$  for misinformation in the sleep before misinformation condition than in the wake before misinformation condition. We found no significant difference between these conditions ( $p = .807$ ). On the other hand, if sleep just before exposure to misinformation benefitted encoding of misinformation, then memory for misinformation should be greater for the sleep before misinformation condition compared to the sleep after misinformation condition. No significant difference was found between these conditions ( $p = .807$ ). Therefore, not only do the misinformation differences between sleep and wake conditions not conceptually replicate, our data do not support either of their explanations.

Our confirmatory analyses seem to indicate that source memory is greater when sleep comes before misinformation compared to when it comes after misinformation. Rather than a post encoding benefit of sleep on memory for misinformation this seems to indicate that sleep is more likely to be involved in pattern separation (i.e., a process whereby the hippocampus forms two independent representations of similar stimuli; see Hanert, Weber, Pedersen, Born & Bartsch, 2017). However, if true, such processes do not provide any benefit over and above a period of wake, as confirmed by the lack of interaction in our exploratory analysis of the source memory data. Our findings instead corroborate findings that sleep benefitted old/new memory but did not play a role in increasing endorsement of misinformation (van Rijn et al., 2017). Future work should acknowledge that evidence that sleep increases the misinformation effect is mixed and that the potential mechanisms are poorly understood.

*No evidence for sleep's role in integrating original information and misinformation*

Mounting evidence indicated that sleep-based memory consolidation may play a role in the integration of new and pre-existing information (e.g., Cairney et al., 2018; Dumay & Gaskell, 2007; Tamminen et al., 2010; Tamminen et al., 2013). Cairney et al. conducted the most similar investigation to our own, finding that sleep increased retrieval induced memory distortion of word locations more so than a period of wake. They noted that memory stabilisation during sleep benefited both the original locations of the words and the distorted memory trace produced by retrieval. The argument was that an optimal memory system should store multiple memory traces, distorted or not to allow for memory updating as predicted by the MTT. Yet, their data also lent themselves to an alternative explanation, that sleep integrated the original and distorted memory traces forming a composite of the two, leading to distorted retrieval. Our data did not support either explanation.

Two outcomes would be expected if either of the two predictions made by Cairney et al. (2018) held in our data. Firstly, if sleep preserves memory traces for accurate and inaccurate information then old/new recognition memory and memory for misinformation would both be greater in the sleep after misinformation condition than in the wake after misinformation condition. Secondly, if sleep is involved in integrating the two conflicting memory traces then source memory for the sleep after misinformation condition should be poorer when compared to the wake after misinformation condition. One of these conditions was met: old/new recognition memory was greater for the sleep after misinformation condition compared to the wake after misinformation condition, conceptually replicating previous research (e.g., Maurer et al., 2015; Payne, Stickgold, Swanberg, & Kessinger, 2008; Hanert et al., 2017; but also see Morgan et al., 2019). However, this does not indicate that stabilization of two memory traces or their integration via sleep leads to memory distortion. Instead, this may indicate that sleep irrespective of the timing of the misinformation preserved memory for only the original information, and not for misinformation.

#### *Limitations of the current study*

Our exploratory analyses suggested that in the test phase (but not in the study phase) participants in the wake before misinformation group reported being sleepier than their sleep condition counterparts. This could have contributed towards recognition memory differences between sleep and wake groups. However, this is unlikely to be a sufficient explanation for any differences given that a larger sleepiness difference was seen between the AM and PM time-of-day controls, and this sleepiness difference did not translate to a difference in recognition memory in any of the memory categories tested. In addition, no sleepiness difference was found between participants in the wake after misinformation group and their sleep counterparts, yet a difference in memory for true events was found between these groups.

*Conclusions*

In our experiment, discriminability between experimental and control conditions did not differ. This may mean that use of the three-phase misinformation paradigm could be problematic. Therefore any analyses reported in this experiment should be treated with caution. Sleep improved old/new recognition memory but did not strengthen misinformation compared to an equivalent period of wake. We found no evidence to support sleep's role in the integration of true information and misinformation. Perhaps standard misinformation paradigms are unsuitable to investigate this research question because it is possible that individuals recognise that the misinformation is untrue when it is presented to them. If this is the case, it suggests the impact of sleep on memory may be selective: only information which is known to be true at the time of encoding might be consolidated, a speculation consistent with findings that sleep selectively consolidates memories that are cued to be remembered at encoding (Saletin, Goldstein, & Walker, 2011). Alternative paradigms where this does not occur such as RID could be more suitable. Ultimately, the mechanisms by which memory for misinformation is encoded, consolidated, and recalled will require further investigation in order to develop methods of mitigating recall of misinformation.



## Chapter 5: Re-Examining the Impact of Sleep on Emotional Memory

### Abstract

Several findings in psychological science have been found to either not replicate or when replicated, the effect sizes are much smaller than the original. This has led researchers to re-examine whether previous findings replicate and determine if they are robust. Not much of this has been done in investigations into the impact that sleep has on memory. One specific topic of investigation is the impact of sleep on *emotional* memory. Some have reported that sleep selectively benefits consolidation of emotional information compared to neutral information while others have failed to observe this selectivity. We argued that the finding that sleep preferentially benefits emotional memory is either not robust and is not replicable, or that these findings can be better explained by memory strength. In our preregistered experiment 120 participants were assigned to a sleep, wake, AM control, or PM control condition. All participants studied negative and neutral images and were later tested on their memory for those images in a recognition memory task. Participants in the sleep condition slept between study and test, and participants in the wake group remained awake during the day. We found that sleep benefits recognition memory overall, but not for negative stimuli over neutral stimuli. We also found that high confidence decisions for lure items were more accurate in the sleep condition. However, time of day effects may account for some of these findings.

## Introduction

It has become increasingly clear that some findings in psychological science fail to replicate (Center for Open Science, 2013). To determine how replicable the findings are, efforts have been focused on recent large-scale collaborations. In one such collaboration, in which many labs tried to directly replicate 100 experiments, only 39% replicated original findings, and effect sizes of those that replicated were much smaller than the original findings (Open Science Collaboration, 2015). One conclusion is that lack of replicability and the smaller effects are indicative of publication bias, where mostly positive results are published, and null results remain in the file drawer (Franco, Malhotra & Simonovits, 2014).

Questionable research practices that have been implicated in the so-called “replication crisis” include 1) stopping data collection early when results become significant, 2) selectively reporting outcomes of analyses that support desired outcomes and not acknowledging the analyses were exploratory, 3) hypothesising after results are known, and 4) conducting underpowered experiments (e.g., John, Loewenstein & Prelec, 2012). To stop these questionable research practices, transparency has been encouraged by advocating pre-registering hypotheses and analysis plans, and conducting power or sensitivity analyses to determine sample sizes prior to collecting data (Munafo et al., 2017). This transparent approach has been adopted by some researchers across sub-disciplines in psychological science, culminating in a more robust basis for scientific progress because reproducibility and transparency is valued.

Whilst many researchers have embraced the recommended practices, others have not. Researchers conducting sleep research, namely sleep-based memory consolidation, have not yet taken on investigating how replicable and robust their findings are. In this type of research, participants are typically asked to memorise a sequence of words, images, or locations before

a period of sleep overnight or wake during the day. Participants are then tested on their memory for those words, images, or locations.

A number of reports published over the past six decades has suggested that a period of sleep, relative to an equivalent period of wakefulness yields greater memory performance on declarative memory tasks (Drosopoulos, Wagner & Born, 2005; Jones, MacKay, Mantua, Schultz & Spencer, 2018; Wilson, Baran, Pace-Schott, Ivry & Spencer, 2012; but see Morgan, Tamminen, Seale-Carlisle & Mickes, 2019). It has been argued that sleep is an optimal brain state for memory consolidation, the process by which newly encoded information is stored and committed into long-term memory (Diekelmann & Born, 2010; Feld & Diekelmann, 2015; Stickgold, 2005). Neural processes that are unique to sleep, such as Slow Wave Sleep (SWS) and sleep spindles, are positively correlated with improved memory performance, and as such are thought to aid consolidation (Diekelmann & Born, 2010; Feld & Diekelmann, 2015; Stickgold, 2005; but see Ackermann et al., 2015; Mantua, 2018; Pan & Rickard, 2015).

The benefits of sleep on memory have been demonstrated using different tasks (e.g., Schönauer, Pawlizki, Köck & Gais, 2013), which has been taken as evidence of a robust phenomenon (e.g., Diekelmann & Born, 2010). Do these findings reflect a true benefit of sleep? To date, there has been one direct replication of the benefit of sleep on motor memory. When accounting for confounds present in the original study the findings did not replicate the original findings (Rickard, Cai, Rieth, Jones & Ard, 2008). Other researchers have levelled scepticism at the reproducibility of findings that sleep benefits problem solving (Hallgató, Győri-Dani, Pekár, Janacsek, & Nemeth, 2013). Recently there have been calls for sleep researchers to increase the sample sizes in an effort to minimise false positives in sleep research (Nemeth et al., 2019). Broadly, there has been an increasing number of researchers questioning the reproducibility of some of these findings (e.g., Lerner & Gluck, 2019.; Morgan et al. 2019; Schönauer, Pawlizki, Köck & Gais, 2013).

One indicator that a finding may not replicate is inconsistent findings across different experiments. For example, the idea that sleep preferentially consolidates negative information over and above neutral information compared to a period of wake has been proposed (Cox et al., 2018). One explanation is that negative information is “tagged” due to the arousal experienced at encoding (Cox et al., 2018; Hu, Stylos-Allan & Walker, 2006; Payne, Stickgold, Swanberg & Kessinger, 2008; Nishida, Pearseall, Buckner & Walker, 2008; Wagner, Gais & Born, 2001). This in turn leads to the selective processing of negative memories during sleep, which leads to greater memory performance that is specific to negative memories and less so for neutral memories (Walker, 2009).

Some experiments reveal a benefit of sleep on emotional information over neutral information and some do not (Cox et al., 2018). In experiments in which sleep did benefit memory for emotional information over neutral information, the experimental designs varied from examining central vs. peripheral memory (Payne et al., 2008), recollection vs. familiarity (Hu, Stylos-Allan & Walker, 2006), sleep deprivation vs. no sleep deprivation (Wagner et al., 2001), and napping vs. no napping (Nishida et al., 2009). However, when employing similar experimental designs other researchers find no selective effect of sleep on emotional memory (e.g., Baran, Pace-Schott, Ericson & Spencer, 2012; Cellini, Torre, Stegagno & Sarlo, 2016; Lehmann, Schreiner, Siefrietz & Rasch, 2016; Lewis, Cairney, Manning & Critchley, 2011). These inconsistencies indicate that the mechanisms underlying the benefits that sleep has on emotional memory are poorly understood and that a replication in its most basic sense is required (i.e., a simple sleep vs wake manipulation using a standard recognition memory task).

One possible explanation for these discrepancies is that sleep does not selectively favour emotional stimuli but is sensitive to memory strength, with sleep selectively consolidating weak or strong memories. For instance, benefits of sleep on memory are more often found in associative memory and free recall tasks, and less often found in recognition

memory tasks (Diekelmann, Wilhelm, and Born, 2009). In free recall tasks, participants are better able to recall lists of words after a period of sleep compared to a period of wake (e.g., Schönauer et al., 2013; but see Schoch, Cordi & Rasch, 2017). The finding that sleep benefits memory in recognition memory tasks is less consistent with some finding benefits on recognition memory (e.g., Hu et al., 2006) and others findings no benefit (e.g., Morgan et al., 2019). The finding that sleep benefits performance on associative memory and free recall tasks (more challenging tasks) and less so on recognition memory tasks (a less challenging task) may reflect a selective benefit of sleep on stronger memories over weaker memories (Schoch et al. 2017).

Another conclusion that has been drawn about the selectivity of sleep-based memory consolidation is that sleep might consolidate the gist of information at the expense of verbatim details. For example, Payne et al. (2009) used the Deese-Roediger-McDermott paradigm (Roediger & McDermott, 1995), where participants studied lists of semantically related words, and after a retention interval that included nap or wake, were tested on recall of those words. Participants were more likely to falsely recall critical lures (i.e., semantically-related words that are not on the study list) after a nap than after a period of wake. Memory for the target words that appeared on lists was equivalent after a nap and wake. Although in another experiment targets words were recalled better after a full night of sleep compared to an equal time of daytime wake, false recall was reduced in the wake group compared to the sleep group. This was taken as evidence to suggest that sleep consolidates gist-based representations of memory, whereby sleep extracts the broader semantic theme of the studied words. Are these gist-based memories weaker memories than verbatim-based memories? Gist-based memories are similar to familiarity-based memories where the content is remembered as having been presented but specific details are not recollected. Likewise, verbatim-based memories are

similar to recollection-based memories in that specified details about the content are remembered.

Some studies found that sleep benefits recollection-based memories and not familiarity-based memories (e.g., Daurrat et al., 2007; Drospolous et al., 2005). In one experiment, participants took part in a standard list learning experiment, had a retention interval and were then tested on their memory for those words (Drospolous et al., 2005). In the retention interval, participants either had a slow wave rich sleep or REM rich sleep. Recollection-based responses were more accurate after a period of slow wave compared to a period of REM sleep.

Given that gist-based memories are similar to familiarity-based and verbatim-based memories involve recollection, this finding does not support the conclusion that during sleep gist-based representations of memories are extracted. To the extent that gist/verbatim and familiarity/recollection based memories, respectively, are related, then it is possible that gist based memories, like familiarity-based memories, tend to be weaker in strength, whereas verbatim-based memories, like recollection-based memories, tend to be greater in memory strength. The point about familiarity-based memories being weaker in strength than recollection-based memories has been long-understood in the basic memory literature (Dunn 2004; 2008). In order to determine if there is a meaningful difference between familiarity and recollection, it is imperative to control for strength, which is typically done by equating for confidence (e.g., Wixted & Mickes, 2010). In the experiments on sleep and memory, however, memory strength has not been considered as a possible confound and therefore not controlled.

Why are emotional stimuli remembered better with sleep? One argument is that neutral memories are weak, like gist- and familiarity-based memories and emotional memories as strong like verbatim- and recollection-based memories. This argument would follow the finding that sleep benefits free recall over recognition and recollection over familiarity.

However, those findings do not support the argument that sleep supports gist-based memories over verbatim-based memories.

Other evidence is conflicting as to the role sleep might play in memory strength. For instance, in Hu et al. (2008) participants studied a series of emotional and neutral stimuli, slept or remained awake, and were tested on their memory for those images. Familiarity-based emotional memories benefitted from sleep, but recollection-based emotional memories did not. This finding held for emotional images but not for neutral images. Comparatively, using the same design, Atienza and Cantero (2008) found that sleep benefitted recollection-based memories of emotional stimuli but that sleep did not have an impact on familiarity-based memories. These conflicting findings call for a replication attempt.

Memory strength was measured in neither experiment. Therefore, it could be case that recollection-based responses reflected strong memories and familiarity-based responses reflected weaker memories (Dunn 2004; 2008). Without considering the potential strength confound the answer is still unknown. Does sleep selectively benefit emotional memories, does sleep benefit strong over weak memories (or vice versa), or once memory is equated on strength, are negative and neutral images impacted equivalently?

Schoch et al. (2017) considered that discrepancy in the findings in the emotional memory and sleep literature may be explained by memory strength. To test this idea, they manipulated the strength of some stimuli presented during encoding in participants in a wake or sleep group. All participants studied negative and neutral images either in the morning or the evening and were tested on those images in a free recall task either in the morning or the evening. Memory strength was increased by asking some of the participants in the wake and sleep groups to immediately recall the images (i.e., following the principles of the testing effect; e.g., Roediger & Karpicke, 2006). There was no overall benefit of sleep compared to wake on

the free recall task. However, sleep benefitted memory for emotional stimuli that had higher encoding strength. They concluded that for sleep to benefit emotional memory, then it had to be sufficiently strong.

Instead of manipulating memory strength at encoding, a participant's confidence can be used as a proxy for memory strength. If emotional memories that are recognised with high confidence are more accurate after sleep than wake, then the conclusions made in Schoch et al. (2017) would be supported. Moreover, the confidence-accuracy relationship is typically very strong (i.e., as confidence increases, accuracy increases), but whether the confidence-accuracy relationship differs with sleep is unknown and may account for some of these differences reported in the literature.

One of the aims of the current experiment was to conceptually replicate the impact that sleep has on emotional memory accuracy (i.e., discriminability). We followed the experimental design implemented by Baran et al. (2012) as closely as possible, but some deviations were necessary, for example polysomnography was not used as this experiment was conducted online. Baran et al. (2012) implemented an AM-PM:PM-AM design, participants were assigned to one of four conditions: Sleep, Wake, AM control, or PM control. During the study phase participants were shown 45 negative and 45 neutral images and were later tested on those images and another 45 negative and 45 neutral images. Participants in the sleep condition had greater discriminability compared to the wake condition and had greater discriminability for negative images compared to neutral images. However in that study there was no benefit of sleep on negative over neutral memories when compared to a period of wake. We hypothesised that discriminability will be higher 1) in the sleep condition vs. the wake condition, 2) for negative stimuli compared to neutral stimuli, and 3) for negative stimuli than neutral stimuli, and the magnitude of this effect will differ between the sleep vs. wake conditions.



Another aim was to measure if sleep differentially affects the confidence-accuracy relationship compared to wake. As this was the first study of its kind, we had no strong predictions about the direction of any effects. We hypothesised that the confidence-accuracy relationship will differ 1) between the sleep condition vs. wake condition, 2) between negative stimuli and neutral stimuli, and 3) for negative stimuli than neutral stimuli, and the magnitude of this effect will differ between the sleep vs. wake conditions

## Methods

The following protocol was pre-registered with the Open Science Framework and is available at <https://osf.io/7uzt8>.

### *Participants*

Sensitivity analysis revealed that by setting the parameters to standard values of  $\alpha = .050$  and  $1 - \beta = 0.80$ ,  $n = 30$  per experimental condition ( $N = 120$ ) can detect an effect size of  $\eta_p^2 = 0.06$ . Participants were randomly assigned to one of two conditions depending on the time of day they participated at, if participants participated in the morning they were assigned to a Wake ( $n = 30$ ) or AM control ( $n = 30$ ) and if they participated in the evening they were assigned to a Sleep ( $n = 30$ ), and PM control ( $n = 30$ ). Unlike in Baran et al. (2012) participants were collected using Amazon Mechanical Turk (mturk) which allows researchers to collect data online. This of course means that the populations between our study and Baran et al. (2012) differ, since they collected from a student population only. The University of California, San Diego (UCSD) Institutional Review Board approved this study (#181525).

Participants were recruited from Amazon Mechanical Turk ([www.mturk.com](http://www.mturk.com)) and were paid \$6.00 for their participation. To ensure good quality data, participation was limited to those who have hit approval rates of at least 85%. Approval rates are based on the quality of the responses on past tasks. Participants were recruited between 8am-11am and 8pm-11pm and

were assigned to one of four conditions: sleep, wake, AM control and PM control. Data collection took place in the United States in states in the Eastern Standard Time zone to ensure that participation only occurred within one time zone. Participants were excluded from data analysis if they scored at or below chance performance on the memory tasks. Data from new participants were collected until we reached  $N = 120$ .

### *Materials*

*Pre-screening questions.* The pre-screening yes/no questions were the following: Are you between the ages of 18-40 years? Are you currently diagnosed with any sleep disorder (e.g. insomnia, sleep apnoea, etc.)? Are you currently diagnosed with any psychiatric disorder (e.g. depression, posttraumatic stress disorder, etc.), or neurological disorder (e.g. mild cognitive impairment, Alzheimer's disease, etc.) which affects memory? Do you work as a shift worker? Have you travelled across time zones within two weeks prior to participating? Are you currently taking any prescribed medication that may affect sleep or memory?

*International Affective Picture System (IAPS).* Following Baran et al. (2012) a stimulus set of 180 images out of 1182 images from the IAPS (Lang, Bradley & Cuthbert, 2008) was selected. The IAPS is a set of pictures with standardised ratings for arousal and valence. Each image ranged in valence from 1 (most unpleasant) to 9 (most pleasant) and arousal from 1 (most arousing) to 9 (least arousing). Baran et al. (2012) selected 90 negative images with a mean valence rating of 2.48 and a mean arousal rating of 5.02 and selected 90 neutral images with a mean valence rating of 5.38 and a mean arousal rating of 2.28. Given that we did not acquire the same images as Baran et al. (2012) we attempted to select images which yielded similar valence and arousal ratings. We selected 90 images that were classified as negative with a mean valence rating of 2.49 ( $SD = .32$ ) and a mean arousal rating of 5.16 ( $SD = .46$ ). We selected 90 images that were classified as neutral with a mean valence rating of 5.09 ( $SD =$

.30) and a mean arousal rating of 2.83 (SD = .35). These ratings were chosen because they were similar to those to in other sleep experiments in which memory for negative and neutral images were compared across sleep and wake conditions (in this case Baran et al., 2012).

Baran et al. (2012) used 60 images as targets and 120 images as lures. Instead we randomly selected Forty-five of the negative images to be targets and 45 to be lures and 45 of the neutral images were randomly selected to be targets and 45 to be lures. This decision was made to keep the base rates for targets and lures equivalent. The images that were negative and neutral and that were targets and lures were the same for every participant. Ninety images were presented during the study phase of this experiment (45 neutral images and 45 negative images intermixed and randomised). During the test phase of the experiment the 90 old images were shown (i.e. 45 emotional and 45 neutral targets presented during the study phase) and 90 new images (i.e. 45 emotional and 45 neutral lures).

#### Sleep-related Questions and Sleep Questionnaires

*Sleep related questions.* Sleep related questions regarded the time that participants went to sleep, woke up, and for how long they slept prior to participating in the study phase.

*Stanford Sleepiness Scale (SSS).* The SSS measures an individual's current sleepiness on a 7-point scale (1 = feeling active, vital, alert or wide awake; 7 = no longer fighting sleep, sleep onset soon; having dream like thoughts) (Herscovitch & Broughton, 1981). Low scores indicate low levels of sleepiness and high scores indicate high levels of sleepiness. The data collected from the SSS were used as a manipulation check to ensure the sleep and wake conditions did not differ on sleepiness at the time of participation. For example, if one condition is sleepier at test compared to another this may impact their performance rather than the manipulation.

*Reduced Morningness Eveningness Questionnaire (rMEQ).* The rMEQ measures chronotype, which is the time of day an individual is most alert (Adan & Almirall, 1991). Participants answered five questions about their sleep habits (e.g. “One hears about “morning” and “evening” types of people. Which ONE of these types do you consider yourself to be?”). Low scores indicate evening types and high scores indicate morning types. Data on this scale were collected to determine whether there are differences in chronotype between conditions. For example, differences between sleep and wake conditions could be due to a greater number of evening types in the sleep condition, who studied the images in the evening. Data from this questionnaire are not analysed in the current chapter.

*Epworth Sleepiness Scale (ESS).* The ESS measures general daytime sleepiness (e.g., would never dose, slight chance of dosing, moderate chance of dosing, high chance of dosing) across several everyday situations (e.g., sitting and reading; Johns, 1991). Higher scores on the ESS indicate greater levels of daytime sleepiness. Data on this scale were collected to rule out differences in general daytime sleepiness as a confound. For example, participants in the wake condition may perform worse because they were sleepier, not because they did not sleep. Data from this questionnaire are not analysed in the current chapter.

*Procedure.* Participants in the AM control condition completed the experiment between 8am-11am and participants in the PM control condition completed the experiment between 8pm-11pm. Participants assigned to sleep and wake conditions completed the experiment in two sessions. In the wake condition participants completed the first part of the experiment between 8am-11am and completed the second part of the experiment between 8pm-11pm. In the sleep condition, participants completed the first part of the experiment between 8pm-11pm and completed the second part of the experiment between 8am and 11am. The time window which participants were allowed to participate in differed from Baran et al. (2012) who collected data between 8pm and 10pm and 8am and 10am as this experiment was conducted

online meaning that participants needed a larger window of opportunity to sign up and take part.

All participants completed the following procedure online: Participants consented, provided demographic information (e.g., gender, education, and ethnicity), and rated their current sleepiness using the Stanford Sleepiness Scale (SSS). Next was the study phase. As in Baran et al. (2012) each trial started with a fixation cross presented for 1500 ms, followed by an image presented for 1000 ms. At their own pace, participants rated the valence (1 = sad, 9 = happy) and arousal (1 = calm, 9 = excited) of each image.

Approximately 12 hours after the study phase for the sleep and wake conditions and immediately after the study phase for the AM and PM control conditions, participants took the SSS. Participants in the wake condition were instructed not to nap between the study and test phases and were asked whether they napped. All participants were asked if they had consumed any caffeine.

Participants then took an incidental (i.e., surprise) recognition memory test following Baran's et al. (2012) design. At their own pace, participants rated the valence (1 = sad, 9 = happy) and arousal (1 = calm, 9 = excited) of each image. Participants were asked, "Have you see this image before?" and responded with a "yes" if they saw the image in the study phase or a "no" if they did not. Unlike in in Baran's et al. (2012) design participants were also asked, "How confident are you?" and responded on a three-point scale (1 = "Guess", 2 = "Sure", 3 = "Very sure"). All decisions were self-paced. Participants completed the rMEQ, ESS, and were asked "What time did you go to bed last night" and "What time did you wake up this morning"? Participants were finally debriefed.

### Analysis Strategy

There were two independent variables. The first was retention interval, sleep vs. wake. The second was the type of stimuli, negative vs. neutral images. There were also two dependent variables. The first,  $d'$ , a measure of discriminability is given by:

$$d' = z(HR) - z(FAR)$$

where hit rate (HR) refers to the number of times participants made a “yes” (i.e., an “old” decision) response to a target divided by the number of target items. False alarm rate (FAR) refers to the number of times participants made a “yes” (i.e., an “old” decision) response to a lure item divided by the number of lure items. This was calculated for negative images, neutral images, and across both types of images. The second dependent variable was proportion correct (i.e.  $H / (H+F)$  for “old” decisions) which was calculated for each participant. We computed a mean over all participants for each of the conditions at each level of confidence (i.e., low, medium and high). Examining the relationship between confidence and proportion correct provides a measure of the confidence-accuracy relationship (i.e. whether confidence and accuracy are related).

The criteria to make inferences about significant differences was  $p < .050$ . Hypotheses where a directional prediction was specified were one-tailed tests, otherwise they were two-tailed tests. No corrections were used for the multiple comparisons for directional predictions, however Bonferroni corrections were applied to any exploratory analyses. Outliers were detected using box-plots of overall  $d'$ . Values more than 1.5 times above or below the third and first quartile range, respectively, were excluded from all analyses and new participants were recruited until the minimum desired sample size was reached.

#### *Confirmatory analyses*

A 2 (sleep vs. wake) x 2 (negative vs. neutral) mixed ANOVA on  $d'$  values was conducted to test the hypotheses related to discriminability. A 2 (retention interval; sleep vs.

wake) x 2 (valence; negative vs. neutral) ANOVA on proportion correct was conducted to test the hypotheses related to the confidence-accuracy relationship.

#### *Preregistered follow-up analyses*

To address whether wake changes emotional reactivity to negative images more so than sleep, we measured the change in valence ( $\Delta$ Valence).  $\Delta$ Valence was calculated by subtracting valence scores given during the test phase from the valence scores given during the study phase. The  $\Delta$ Valence for emotional images has been reported as greater for the wake condition compared to the sleep condition (i.e. images were rated more positively after a period of wake; e.g., Baran et al., 2012). A 2 (sleep vs. wake) x 2 (negative vs. neutral) ANOVA with change in valence as the outcome measure was conducted. The ANOVA was conducted to reveal any main effects of sleep vs. wake, negative vs. neutral stimuli, and interaction between sleep vs. wake and type of stimuli on  $\Delta$ Valence.

To rule out time-of-day effects we compared  $d'$  between the two (circadian) control conditions. A 2 (AM control vs. PM control) x 2 (negative vs. neutral) ANOVA on  $d'$  was conducted to test differences in discriminability. The ANOVA was conducted to reveal any main effects of AM control vs. PM control, emotion vs. neutral, and interaction between AM control/PM control and type of stimuli on discriminability.

The confirmatory analyses outlined above were conducted based on 1) the pre-defined norms for which targets and lures are negative and which are as specified in IAPS and 2) the valence ratings participants provided for targets during the study phase and the lures during the test phase. Ratings  $>5$  were classified as negative and ratings  $\leq 5$  were classified as neutral.

## Results

Three participants performed below chance and two participants were outliers and therefore excluded from the analysis and five more participants were recruited as per our analysis strategy. Table 14 shows demographic information of the sample ( $N = 120$ ). Of the entire sample, 21 reported having a sleep disorder,  $n = 6$  having a neurological disorder,  $n = 16$  reported having a psychiatric disorder,  $n = 11$  reported taking medication that affects sleep, no-one reported taking medication that affects their memory,  $n = 10$  reported recently travelling across time zones,  $n = 9$  reported that they currently work as a shift worker, and  $n = 3$  in the sleep condition reported sleeping less than 6 hours ( $M = 6.79$ ,  $SD = 1.46$ ). None of these participants were excluded on the basis of these criteria given that this would reduce our power to detect the effect size of interest reported in our power analysis. The results did not vary depending on whether the analyses were conducted using the predefined norms from IAPS or participants ratings of the images, so the latter are reported.

**Table 14.** Demographic characteristics per condition.

	Condition			
	Sleep	Wake	AM Control	PM Control
Gender				
Female	17	14	19	7
Male	13	16	11	23
Age				
Years ( <i>SD</i> )	37.17 (11.03)	39.93 (10.48)	38.87 (11.74)	31.10 (5.78)
Ethnicity				
African American	2	0	2	3
Arab American	0	0	0	1
Asian American	0	2	1	0
Caucasian American	25	25	27	23
Education				
Bachelor's Degree	11	13	14	13
High School/GED	3	3	6	6
Master's Degree	3	3	2	1
Post-Masters	0	0	3	3
Some College	11	11	7	7
Do not wish to state	1	0	0	0

Mean,  $M$ ; Standard deviation,  $SD$ ; Confidence intervals,  $CI$ .



*Confirmatory analyses**Discriminability*

*Sleep vs. wake.* Figure 8A shows that  $d'$  is higher for participants in the sleep condition compared to participants in the wake condition,  $F(1, 58) = 5.03$ ,  $p = .030$ ,  $\eta_p^2 = .080$ , 90% CI [0.00, 0.20]. There was no significant main effect of valence on  $d'$ ,  $F(1, 58) = 1.26$ ,  $p = .266$ ,  $\eta_p^2 = .021$ , 90% CI [0.00, 0.11], and no significant interaction between sleep vs. wake and negative vs. neutral images on  $d'$ ,  $F(1, 58) = 0.47$ ,  $p = .494$ ,  $\eta_p^2 = .008$ , 90% CI [0.00, 0.08]. Means and standard deviations are presented in Table 15.

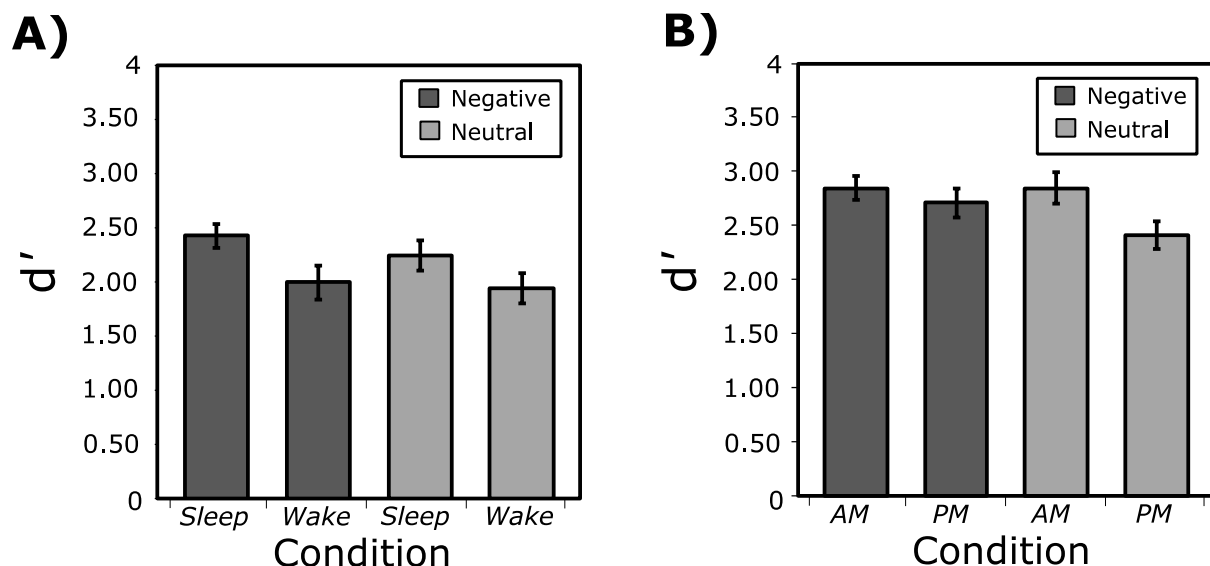


Figure 8. Mean  $d'$  across negative and neutral items for sleep and wake conditions (A) and AM and PM conditions (B). Error bars represent standard error of the mean.

**Table 15.** Mean  $d'$ , standard deviation, and 95% confidence intervals for negative and neutral stimuli per experimental condition.

Valence		<i>M</i>	<i>SD</i>	95% CI
<i>Negative</i>		2.21	0.71	[2.03, 2.38]
	Sleep	2.43	0.61	[2.18, 2.68]
	Wake	1.98	0.75	[1.73, 2.24]
<i>Neutral</i>		2.10	0.81	[1.89, 2.31]
	Sleep	2.26	0.84	[1.97, 2.56]
	Wake	1.95	0.76	[1.65, 2.24]
<i>Overall</i>				
	Sleep	2.35	0.74	[2.11, 2.58]
	Wake	1.96	0.75	[1.72, 2.21]

Mean,  $M$ ; Standard deviation,  $SD$ ; Confidence intervals, CI.

### *Confidence-Accuracy Relationship*

We did not have enough data in each confidence category to analyse these data as planned (i.e., low, medium and high confidence categories). This is because some participants were less likely to choose the low and medium confidence ratings for “old” decisions. Therefore, in a necessary deviation to our preregistered analyses we only conducted ANOVAs on proportion correct to high confidence responses for old decisions, where there was enough data for each participant.

A 2 x 2 mixed ANOVA on proportion correct for high confidence responses was conducted. There was no significant main effect of sleep vs. wake on high confidence accuracy,  $F(1, 58) = 0.85$ ,  $p = .359$ ,  $\eta_p^2 = .015$ , 90% CI [0.00, 0.10]. There was also no significant main effect of valence on high confidence accuracy,  $F(1, 58) = 0.02$ ,  $p = .961$ ,  $\eta_p^2 < .001$ , 90% CI [0.00, 0.03], nor was there an interaction between sleep vs. wake and negative vs. neutral,  $F(1, 58) = 3.34$ ,  $p = .065$ ,  $\eta_p^2 = .006$ , 90% CI [0.00, 0.17].

To examine any time of day differences in high confidence accuracy a further 2 x 2 mixed ANOVA was conducted. There was no significant main effect of time of day on high confidence accuracy,  $F(1, 58) = .001$ ,  $p = .980$ ,  $\eta_p^2 < .001$ , 90% CI [0.00, 0.00]. There was also

no significant main effect of valence on high confidence accuracy,  $F(1, 58) = .007$ ,  $p = .932$ ,  $\eta_p^2 < .001$ , 90% CI [0.00, 0.01], nor was there an interaction between time of day and negative vs. neutral,  $F(1, 58) = 0.79$ ,  $p = .376$ ,  $\eta_p^2 = .014$ , 90% CI [0.00, 0.09].

#### *Preregistered follow up analyses*

*ΔValence.* Given that previous literature has proposed that sleep impacts the emotional reactivity of images, we performed a 2 (sleep vs. wake) x 2 (negative vs. neutral) ANOVA with change in valence as the dependent variable. A positive value indicates that the valence became more positive and a negative value indicates that an image became less positive. The ANOVA revealed no significant main effect of sleep vs. wake on change in valence,  $F(1, 58) = 1.73$ ,  $p = .194$ ,  $\eta_p^2 = .029$ , 90% CI [0.00, 0.13]. There was a significant main effect of negative vs. neutral stimuli on change in valence,  $F(1, 58) = 77.69$ ,  $p < .001$ ,  $\eta_p^2 = .573$ , 90% CI [0.42, 0.66]. Negative images became more positive ( $M = 0.42$ ,  $SD = 0.50$ ) compared to neutral images that became less positive ( $M = -0.35$ ,  $SD = 0.44$ ). There was no significant interaction between sleep vs. wake and negative vs. neutral stimuli on change in valence,  $F(1, 58) = 0.67$ ,  $p = .797$ ,  $\eta_p^2 = .001$ , 90% CI [0.00, 0.09].

*AM control vs. PM control.* Figure 8B shows average  $d'$  for negative and neutral images for the AM and PM conditions. A 2 x 2 mixed ANOVA revealed no significant main effect of time of day on  $d'$ ,  $F(1, 58) = 3.12$ ,  $p = .083$ ,  $\eta_p^2 = .051$ , 90%, CIs [0.00, 0.16], no significant main effect of negative vs. neutral on  $d'$ ,  $F(1, 58) = 3.06$ ,  $p = .086$ ,  $\eta_p^2 = .050$ , 90% CIs [0.00, 0.16], and no significant interaction between time of day and negative vs. neutral,  $F(1, 58) = 3.04$ ,  $p = .087$ ,  $\eta_p^2 = .050$ , 90% CIs [0.00, 0.16]. Means and standard deviations are presented in Table 3.

#### *Exploratory analyses*

*Experimental vs. control.* As there was a 12-hour retention interval between study and test phases for the sleep and wake conditions and no retention interval for AM and PM control conditions, memory for the control conditions should be greater than the experimental conditions. An independent *t*-test revealed that recognition memory for the control conditions ( $M = 2.62$ ,  $SD = 0.65$ ) was greater than recognition memory for the experimental conditions ( $M = 2.08$ ,  $SD = 0.68$ ),  $t(118) = 4.36$ ,  $p < .001$ ,  $d = 0.81$ , 95% CIs [0.44, 1.18].

*Sleepiness.* Any benefits on recognition memory of sleep could be explained by differences in sleepiness at study and test between sleep and wake conditions. The sleep condition reported being sleepier ( $M = 2.63$ ,  $SD = 1.33$ ) than the wake condition ( $M = 2.03$ ,  $SD = 1.22$ ) at study, but the difference was not significant,  $t(58) = 1.83$ ,  $p = .073$ ,  $d = 0.47$ , 95% CIs [-0.05, 0.98]. There were also no differences in sleepiness at test between sleep ( $M = 2.70$ ,  $SD = 1.53$ ) and wake ( $M = 2.40$ ,  $SD = 1.52$ ) conditions,  $t(58) = 0.76$ ,  $p = .450$ ,  $d = 0.19$ , 95% CIs [-0.31, 0.70]. Sleepiness for AM and PM control conditions at study and test was also examined, there were no differences between AM ( $M = 1.96$ ,  $SD = 0.93$ ) and PM ( $M = 2.26$ ,  $SD = 1.34$ ) controls at study,  $t(58) = 1.01$ ,  $p = .317$ ,  $d = 0.26$ , 95% CIs [-0.25, 0.76]. However sleepiness was significantly higher in the AM control condition ( $M = 1.83$ ,  $SD = 0.91$ ) compared to the PM control condition ( $M = 2.53$ ,  $SD = 1.46$ ) at test,  $t(58) = 2.23$ ,  $p = .029$ ,  $d = 0.57$ , 95% CIs [-0.05, 1.08].

*Confidence-Accuracy Relationship.* Proportion correct was also calculated for “new” responses (i.e. CR/CR+M)) to determine whether proportion correct for “new” decisions varied for sleep and wake conditions and for emotional and negative stimuli at different levels of confidence (i.e., low, medium and high). “New” decisions refer to “no” responses to targets and lures. If a participant responds “no” to a target, that is a miss (M), if a participant responds “no” to a lure, then that is a correct rejection (CR).

Due to the number of responses made in the low and medium confidence levels, we were unable to conduct ANOVAs for proportion correct for these confidence levels for “old” decisions, as planned in our pre-registration. This was also the case for “new” decisions. One benefit of calibration plots is that they are a graphic representation of the confidence-accuracy relationship. Calibration plots are created by plotting proportion correct for “old” and “new” responses against each level of confidence. In Figure 9 proportion correct is pooled across participants instead of calculating it individually for each participant and allow for a graphical analysis of “old” and “new” decisions. This means that there is enough data to calculate proportion correct for each level of confidence for “old” and “new” decisions. In Figure 9, overlapping error bars signify non-significant differences (Mickes, 2015).

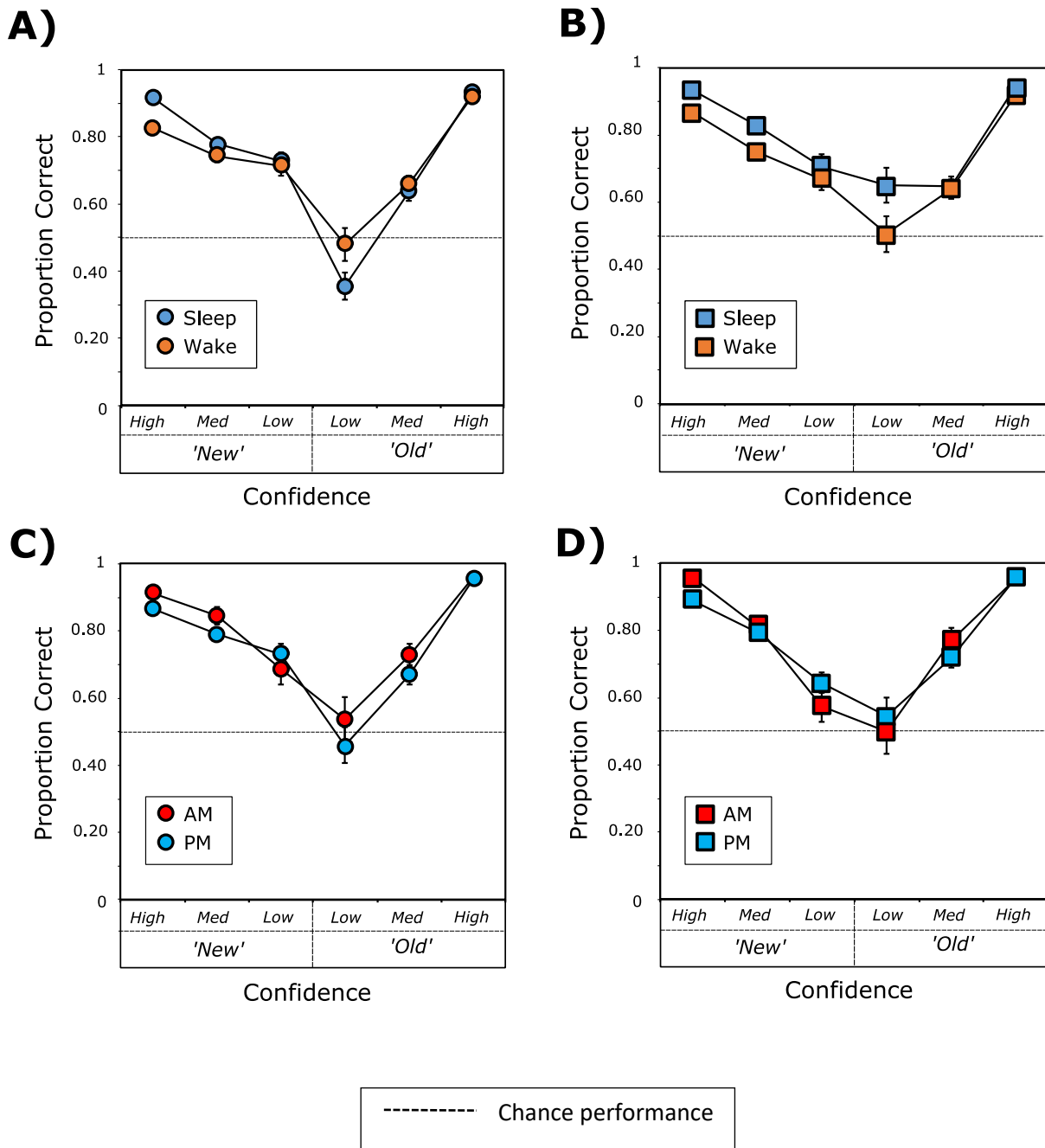


Figure 9. Calibration plots for neutral images (A) and negative images (B) in sleep and wake conditions. C) Calibration curves for neutral images (C) and negative images (D) in AM and PM control conditions. Error bars represent standard error of the mean.

In all of the conditions, for negative and neutral images, the proportion correct across the levels of confidence followed similar patterns. Confidence and accuracy were related: High confidence responses were more accurate than medium confidence responses and medium

confidence responses were more accurate than low confidence responses. In the sleep and wake conditions responses to old items (i.e., targets) made with medium and high levels of confidence had similar performance. However, the sleep condition had higher levels of performance compared to the wake condition for new items (i.e., lures) for responses made with medium and high levels of confidence. That pattern was the same for the AM control condition and PM control condition (i.e., similar proportion correct for targets, but the AM control condition had higher proportion correct for new items than PM control condition). The negative vs. neutral images did not appear to have an effect on proportion correct for any of the conditions.

## Discussion

Sometimes sleep, compared to a period of wake, selectively increases memory for negative stimuli over neutral stimuli (Cox et al., 2018; Hu, Stylos-Allan & Walker, 2008; Payne, Stickgold, Swanberg & Kessinger, 2008; Nishida, Pearseall, Buckner & Walker, 2009; Wagner, Gais & Born, 2001). One proposed explanation is that emotional memories are “tagged” so that the consolidation process that occurs during sleep selectively benefits emotional stimuli (e.g., Payne et al., 2008). Neutral stimuli should not receive the same benefits to the same magnitude. However, sometimes sleep does not selectively increase memory for negative stimuli. These conflicting findings may arise because of various differences in experimental design. We therefore conducted a well-powered experiment in a conceptual replication attempt following Baran's et al (2012) experimental design. We found that sleep increased discriminability for both types of stimuli, but did not selectively improve memory for negative images over and above memory for neutral images compared to a period of wake. However, we also did not find that negative images were more memorable than neutral images, which limits conclusions that can be made.

We had also proposed that the discrepancies across findings may be because sleep selectively effects strong or weak memories. To measure this, we collected confidence and presented calibration plots. Participants in all conditions had strong confidence-accuracy relationships, and participants in the sleep condition had higher proportion correct when they gave high confidence responses to new items. However, time-of-day effects limit conclusions that sleep improves memory.

*The Impact of Sleep on Emotional Memory: Discriminability*

Our findings support the claims that sleep benefits recognition memory overall (Baran, Pace-Schott, Ericson & Spencer, 2012; Cellini, Torre, Stegagno & Sarlo, 2016; Lehmann, Siefrietz & Rasch, 2016; Lewis, Cairney, Manning & Critchley, 2011). However, our finding that sleep did not selectively increase memory for negative information over and above neutral information conceptually replicates some findings (Cellini, Torre, Stegagno & Sarlo, 2016; Lehmann, Schreiner, Siefrietz & Rasch, 2016; Lewis, Cairney, Manning & Critchley, 2011). Although sleep appears to benefit overall recognition memory, we also found an effect of time of day (the difference was not significant, but the effect size medium). Participants in the AM control condition had greater recognition memory compared to participant in the PM control condition. Both sets of participants in the AM and Sleep conditions were tested in the morning and those in the PM and Wake conditions were tested in the evening. Therefore, the benefit of sleep on recognition memory might be because participants were tested in the morning and not the evening.

The finding that emotional stimuli are more memorable than neutral stimuli is a robust one (e.g., Cox et al., 2018; Hu, Stylos-Allan & Walker, 2006; Payne, Stickgold, Swanberg & Kessinger, 2008; Nishida, Pearseall, Buckner & Walker, 2009; Wagner, Gais & Born, 2001). However, we did not get a differential effect, which may be why we did not find a selective



benefit of sleep on memory for negative stimuli. If the “tagging” hypothesis of emotional memories is correct, then observing selective sleep benefits for emotional memories is likely contingent on observing a benefit of negative images over neutral images. It is reasonable to assume that it is essential for negative images to have greater memory in comparison to neutral images for the benefit of sleep on emotional memory to emerge.

*The Impact of Sleep on Emotional Memory: Confidence-Accuracy Relationship*

We proposed that the discrepancies in previous findings examining the relationship between sleep and emotional memory could be explained by varying levels of memory strength. That is, whether sleep benefits strong over weak memories (or vice versa) is an unanswered question. We had no strong predictions about the impact of memory strength on the selectivity of sleep as research was mixed and indicated that sleep could benefit weak memories (i.e., gist or familiarity-based memories; Daurrat et al., 2007; Drospolous et al., 2005; Hu et al., 2008; Payne et al., 2009) or strong memories (i.e., verbatim or recollection-based memories; Attienza & Cantero, 2008; Schoch et al., 2017).

The confidence-accuracy relationship was similar for negative and neutral stimuli. The confidence-accuracy relationship was also similar for “old” decisions across all of the conditions. Interestingly, sleep does appear to benefit stronger “new” memory decisions made with high confidence. Thus, participants who slept between study and test were better at correctly rejecting images that they did not see during the study phase compared to the wake condition. However, a similar pattern occurred in the time of day controls. That is, there was a significant difference in high confidence “new” decisions in the AM control condition than for the PM control condition, but not for the medium or low confidence levels. One possibility is that encoding the stimuli in the morning somehow enables participants to better correctly reject

stimuli that were not shown during the study phase. As this is some of the first analyses of this kind, and is exploratory, in sleep research more experiments are needed.

### *Limitations*

It is possible that the inconsistencies in findings also reflect a much broader issue of lack of power in sleep research. Recently Nemeth, Gerbier and Janacsek (2019) acknowledged that sample sizes in sleep research range around 40 participants per experiment, with anywhere between 12-20 participants per condition; though experiments with larger sample sizes have been published (e.g., Fenn & Hambrick, 2012). In our experiment, we were powered to detect a maximum of a medium effect size,  $\eta_p^2 = .060$ , with 30 participants per condition. It is possible that the effect is smaller than we could detect despite the fact that researchers have reported large effects, which we theoretically should be powered enough to detect with our sample (e.g. Hu et al., 2006,  $\eta_p^2 = .200$ ).

We did not find a benefit of sleep on emotional memory relative to neutral memory, despite being powered to detect an effect much smaller than has been found previously. Of course, however, one could argue that our definition of replication is unsuitable, that is  $p < .050$ . Indeed, other means of determining replication have been proposed including determining whether the effect size falls within the confidence intervals of effect sizes found previously (Rahal & Open Science Collaboration, 2015). However, in many of the papers published the necessary information (effect sizes and confidence intervals around the effect size) are not reported (e.g., Cox et al., 2018). This issue has been faced before and calls for more open sharing of data within sleep research have been made (Nemeth, Gerbier & Janacsek, 2019). Such issues make it difficult to define replication within sleep research which can slow down the progress of our theoretical understanding of the function sleep plays in consolidation.

Nevertheless, effect size and confidence intervals can be estimated for an ANOVA using the  $F$ -value, numerator df, denominator df and the sample size of a given study. We used this information to calculate the confidence intervals around the effect size found in the most recent investigation of the impact of sleep vs. wake on emotional memory conducted by Cox et al. (2018), specifically their interaction effect. Cox et al. reported a small to medium effect,  $\eta_p^2 = .077$  95% CIs: .006, 0.189, which means that we should also be able to detect an effect. Our effect size was  $\eta_p^2 = .008$  95% CI: .000, 0.080, and so our effect does fall within the confidence intervals of the Cox et al. study, which could be considered a conceptual replication albeit smaller than the aforementioned study. However, our effect is very small and unlikely reflects a robust difference in memory for emotional vs. neutral images in the sleep condition. Given that the effect found by Cox et al. ranges from miniscule to medium effect also indicates that perhaps the benefits of sleep on emotional memory are not robust, and that larger sample sizes are required to detect the true effect. Regardless, any interpretation of our findings should be interpreted with caution given the potential time of day effects present in our experiment.

### *Conclusion*

We found that sleep benefits discriminability irrespective of emotional valence, but we did not conceptually replicate the previously reported finding that sleep preferentially consolidates negative stimuli over neutral stimuli compared to wake. However, this may be because memory for negative stimuli was not better than memory for neutral stimuli. We were unable to directly replicate any of the relevant past experiments because there was insufficient information reported to do so.

We proposed that the inconsistencies in previous research could be due to the variable impact of sleep at different levels of memory strength. To test this, we used participants' confidence as a proxy for memory strength (i.e. accuracy), otherwise known as the confidence-

accuracy relationship. In an exploratory analysis no effect of sleep compared to wake was found on memory performance at different levels of confidence for “old” decisions, but differences were found for high confidence “new” decisions across negative and neutral stimuli, and for medium confidence decisions in negative stimuli. However, it is unclear whether sleep is solely responsible for this difference, as time of day also appears to yield a similar if not identical pattern of results. Regardless, further research is required to determine whether sleep impacts emotional memory processing at varying levels of memory strength.

## Chapter 6: General Discussion

The main aim of this thesis was to examine the impact of sleep on recognition memory. The hypotheses were based on findings in the scientific literature that sleep strengthens memory or at least provides a passive protection from interference that occurs during waking. There is a commonly held view that sleep benefits different types of memory, and specific stages of sleep and neurological activity during sleep facilitate those benefits. In other words, there seems to be a widespread benefit of sleep on memory that is supported by a large number of studies that show that participants who sleep after learning compared to those who remain awake perform better on memory tasks (e.g., Brawn, Fenn, Nusbaum & Margoliash, 2010; Fischer et al., 2005; Korman et al., 2003, 2007; Schönauer, Pawliziki, Köck & Gais, 2014; Walker, Brakenfield, Morgan, Hobson & Stickgold, 2002; Walker, Stickgold, Alsop, Gaab & Schlaug, 2005; Drosopoulos, Schulze, Fischer & Born, 2007; Payne et al., 2012; Tucker & Fishbein, 2008; Wilhelm et al., 2011; Tucker, Tang, Uzoh, Morgan & Stickgold, 2011; Drosopoulos, Wagner & Born, 2005; Wagner, Kashyap, Diekmann & Born, 2007).

While there are many experiments that show sleep benefits memory, there are only a few experiments on more applied matters. The two first experiments were designed to investigate the impact of sleep on eyewitness identifications and susceptibility to false memories. We found difference in memory performance between sleep and wake conditions, so a third experiment, aimed at conceptually replicating the benefits of sleep on memory for emotional stimuli was conducted. The results across the experiments did not tell a consistent story.

Table 16 shows a summary of the experiments, sample and trial sizes, and findings. The data collected in the eyewitness identification experiment (presented in Chapter 2) were used to provide normative data of two commonly used sleepiness scales and sleep duration that can be used by clinicians and researchers about sleep health. There were no meaningful differences

in sleepiness or sleep duration across gender, age, and education, but there was a difference that reveals a potential confound in sleep experiments. Below the results are discussed in more detail alongside theoretical considerations, including explorations of potential reasons for the inconsistent findings and potential solutions and future directions.

**Table 16.** Sample size, sleep benefit,  $p$  value, effect size, and sleepiness and time of day confounds for each experiment. These values are from the pre-registered analyses for comparisons between sleep and wake groups.

Experiment	Sample Size	No. trials per participant	Sleep Benefit	$p$ value	Effect Size	Sleepiness Confound	Time of Day Confound
Chapter 2: Eyewitness identification	4,000	1	No	.892	$pAUC_{dif} < .001$	Yes	No
Chapter 3: Norms	7,533	-	-	-	-	-	-
Chapter 4: False memories	300	72	?	.047	$F = .016$	No	No
Chapter 5: Memory for emotional stimuli	120	180	Yes	.030	$\eta_p^2 = .080$	No	Yes

## Summary of findings and Theoretical Considerations

### *Sleep based memory consolidation and recognition memory*

*Examining the impact of sleep on eyewitness identifications.* In Chapter 2, an experiment on the impact that sleep has on eyewitness identification performance, specifically discriminability (i.e., the ability to discriminate between innocent vs guilty suspects) and reliability (i.e., the likelihood that the identified suspected is guilty), was investigated. The hypothesis that sleep would lead to greater discriminability compared to an equivalent period of wake based on existing literature indicating the sleep yields greater recognition memory performance compared to a period of wake was not supported. No strong predictions were made about reliability because this was the first investigation of its kind. Sleep also did not

impact reliability, but confidence and accuracy were related, and similarly for the sleep and wake conditions.

Although there were no differences in discriminability and reliability, this does not imply that sleep did not passively or actively preserve memory for the perpetrator, it may be that consolidation occurring during wakefulness in the wake condition was also sufficient to preserve lineup memory (Mednick et al., 2011; Wamsley, 2019). Researchers are finding that wake in addition to sleep is another state in which consolidation can occur (Mednick et al., 2011). For instance, some researchers have found that wakeful rest can result in consolidation which resembles the benefits shown during sleep on declarative memory tasks (e.g., Foster & Wilson, 2006). Wamsley et al. propose that the cellular consolidation which occurs immediately after learning during wake may be sufficient for subsequent reactivation during resting wakeful periods, as it does during sleep. This occurs because such process can be facilitated during periods where individuals are awake but less engaged in cognitive and attentional processes (Wamsley, 2019).

Of course, we have no way of telling whether participants in the wake condition underwent wakeful rest, but it does remain a plausible explanation for why no difference in discriminability was found. In other words, no difference in between sleep and wake conditions was found because consolidation occurred during both sleep and wake, and this was sufficient to equally preserve lineup memory for a 12-hour period. Additionally determining sleep's role in lineup memory, if any, proved to be difficult since we do not have PSG measures to determine whether sleep specific activity is related in any way to lineup memory performance. This is a shame since studies published since beginning our registered report appear to indicate that activity during NREM and REM sleep play an intricate role in consolidation of faces (Solomonova et al., 2017). This may suggest that sleep in some way does facilitate the active consolidation of faces, although other researchers have argued that the role of sleep in face

recognition may be passive (Sheth et al., 2009). Regardless this experiment indicates that eyewitness identifications may not be positively or negatively impacted by a period of wake or sleep over a 12-hour period. Future research may seek to examine the effects of sleep on eyewitness identifications over longer periods of time, where the consolidative processes such as systems level consolidation have more time to unfold (e.g., Igloi, Gaggioni, Sterpenich & Schwartz, 2015).

*Examining the impact of sleep on the misinformation effect.* In Chapter 4, the impact of sleep on susceptibility to false memories was investigated. The predictions were based on when misinformation was presented (before or after sleep) to investigate whether sleep decreases or increases the likelihood that misinformation is endorsed. The former would be expected when misinformation is presented after sleep and the latter would be expected when misinformation is presented before sleep. The hypothesis that sleep would improve old/new recognition memory was supported, but sleep did not impact memory for misinformation or source memory.

Moreover, it was also predicted that presenting misinformation after sleep when compared to wake would increase the misinformation effect (i.e., attributing inaccurate information to another source; in this case attributing misinformation to the original event). Calvillo et al. (2016) proposed two theoretical possibilities that led to the increase in the misinformation effect when misinformation is presented after a period of sleep. The first proposal was that sleep, via systems level consolidation, extracted the gist of the original event at the expense of verbatim details, and therefore increased endorsement of misinformation (Calvillo et al., 2016). The second proposal was that sleep enables the improved encoding of new information after sleep, via synaptic homeostasis. In other words, the downscaling of memory representations at the synapses reduces LTP within neurons and therefore allows for subsequent LTP after sleep during subsequent wake (Tononi & Cirelli, 2003). Neither of these



theoretical possibilities held in our conceptual replication of their experiment. We conceptualised this in two ways, one by measuring memory for misinformation and two by measuring source confusion that is inaccurately attributing memory for the misinformation to the original event. Neither memory for misinformation, nor source memory differed between the wake before misinformation and sleep before misinformation conditions.

The answer to this failed conceptual replication attempt could be twofold – as was discussed regarding the findings in Chapter 2. That is, two reasons may explain the absence of a difference between these two conditions. The, first, the systems consolidation hypothesis would also posit that sleep should reduce the misinformation effect because systems consolidation during sleep would render memory for the original event more resistant to interference. The second, would posit that the absence of interference with the original memory trace occurring immediately after encoding the original memory would allow for at least LTP and perhaps consolidation during waking periods so that the memory would become stabilised (Mednick et al., 2011). Since the latter may be the explanation for an absence of a sleep benefit on line-up memory it may also explain the findings presented in this chapter, especially since there was also no difference in old/new recognition memory performance between these conditions. In other words consolidation of the original event occurred to an equivalent extent across these sleep and wake conditions and therefore no difference in the endorsement of misinformation or memory for the original event were identified – at least within a 12-hour period.

Based on prior research we also expected that sleep may be involved integrating the original event and misinformation. This prediction was largely driven by two lines of evidence: 1) research indicating that systems consolidation is involved in integrating linguistic information into pre-existing networks of knowledge (Tamminen et al., 2010); and 2) that retrieval of an initial memory trace yields a second memory trace which is preserved during

sleep, and may in this case lead to memory distortion (Cairney et al., 2018). If this was the case, three outcomes would be expected, memory for the original event and the misinformation should be greater in the sleep condition compared to the wake condition and source memory may be worse in the sleep condition compared to the wake condition. Only one outcome held, memory for the original event was greater in the sleep after misinformation condition compared to the wake after misinformation condition. Therefore, it seems under these circumstance that sleep did not increase the misinformation effect as we predicted and instead preserved memory for the original event.

This finding may reflect an active selective component of sleep in memory consolidation, whereby participants may have identified that the misinformation was false, and that this, perhaps via "tagging", enabled the original event to be selectively consolidated and not the misinformation. Indeed, evidence has indicated that when participants are informed that their memory will be tested for information that they have just studied prior to a period of sleep or a period of wake, their memory for that information is greater after sleep (e.g., Wilhelm et al., 2011). The mechanism by which this may have occurred in our experiment is unclear, but neuroscientific evidence could point towards a plausible explanation (Born & Wilhelm, 2012).

In my experiment, participants were informed that their memory would be tested and therefore it is probable that prefrontal brain regions activated in anticipation of a memory test in order to accommodate the recently encoded information in the form of tagging to match the participants' motivation to remember the original event (see Born & Wilhelm, 2012). Specifically participants may have been motivated to remember the original information and not the misinformation when it is identified. Indeed it would be adaptive for participants to forget the misinformation which they identify to reduce the overlap in the representation of the original event (e.g., Anderson & Green, 2001). In a review, Born and Wilhelm (2012) described research which has demonstrated that replay occurs in both the hippocampus and prefrontal

regions during sleep (e.g., Peyrache, Khamassi, Benchenane, Wiener, & Battaglia 2009) and that some of the strongest slow oscillation activity originates from the prefrontal cortex (e.g., Massimini, Huber, Ferrarelli, Hill, & Tononi 2004). Therefore it has been speculated that the interaction between the hippocampus and prefrontal cortex during learning of relevant information tags those memories and the brain regions which represent that memory are subsequently reactivated during sleep and which in turn stabilises those relevant memories (e.g., the original event) and not those which are not relevant (e.g. misinformation). Indeed, evidence has demonstrated that sleep can drive the remembrance of "to be remembered information" and forgetting of "to be forgotten information" (e.g., Saletin, Goldstein & Walker, 2011). This would allow for the selective enhancement of memories for the original event which are motivationally relevant to the participant and not the misinformation which should be forgotten.

However we cannot confirm whether participants identified the misinformation because they were not asked, and memory for misinformation ( $M = 1.13$ ,  $SD = 0.66$ ) in both sleep and wake after misinformation conditions was higher than that of memory for the original event ( $M = 0.96$ ,  $SD = 0.47$ ;  $F(1, 98) = 5.99$ ,  $p = .016$ ,  $\eta_p^2 = .058$ , 90% CI [0.01, 0.14]). Whilst it is possible that sleep may have reduced the magnitude of difference in memory performance between old new recognition memory and memory for misinformation via tagging of the original event, this did not result in an interaction  $F(1, 98) = 0.85$ ,  $p = .771$ ,  $\eta_p^2 = .001$ , 90% CI [0.00, 0.20]). Therefore it is unclear why memory for the original event is improved after a period of sleep in this instance, and perhaps has occurred by some other mechanism which we do not currently understand but one that of course should be explored. Importantly future research should ask participants if they did identify the misinformation and if this is related to selective retention of true information and forgetting of misinformation.

One puzzling finding was that the experimental and control conditions had similar  $d'$  values despite the 12-hour difference in retention intervals. Given this difference, the control conditions should have had much higher  $d'$  values. Similar findings have been reported in other experiments in which very similar paradigms were used (e.g., Lo et al., 2016). These findings may signal a fundamental flaw in the paradigm and present a major challenge to making interpretations of the impact that sleep has on memory. Preliminarily however, we can conclude on the basis of the findings in Chapter 2 and Chapter 4 that sleep indeed does not harm eyewitness memory, does not benefit line-up memory but may benefit episodic memory for crimes when compared to a period of wake.

*Sleep and emotional memory: A conceptual replication attempt.* Given that there was no strong evidence that sleep benefits recognition memory, I attempted a conceptual replication of previous research (presented in Chapter 5). In this experiment, the impact of sleep on emotional memory, where sleep is predicted to benefit emotional memories more so than neutral memories compared to a period of wake, was investigated. The evidence demonstrating this effect previously was mixed. Therefore, a new theoretical mechanism that could explain the inconsistencies in previous findings, namely the influence sleep has on memory strength, was investigated. I had no strong predictions regarding memory strength or the interaction between sleep and emotional memory, but I did predict that sleep would benefit memory compared to a period of wake and that negative images would be better recognised compared to neutral images. There was no evidence for an interaction between sleep and valence. However, sleep yielded greater recognition memory compared to a period of wake and sleep benefitted high confidence “new” decisions. The control conditions yielded similar patterns (though not significant, the effect sizes were medium), therefore differences between sleep and wake may be confounded by time of day effects. It is unclear whether the results reflect a benefit of sleep on memory or a circadian effect on memory.

As I acknowledged in Chapter 5 we were unable to find a difference in memory performance for negative vs. neutral images. It is plausible that this led to absence of an interaction between sleep and emotional memory. For example, in order for tagging of emotional information to occur it is predicted that an emotional image should evoke activation in the amygdala which leads to synaptic plasticity across multiple brain regions including the hippocampus (see Hutchison & Rathore, 2015 for a review; van der Helm & Walker, 2009). The activation of the amygdala is also proposed to cause strengthening at the synapses which represent the emotional images thus leading to long term consolidation of those images. Since there was no difference in memory for neutral vs. negative images it is possible that the negative images did not evoke a significant activation within the amygdala to result in the tagging necessary for an overall benefit on memory for emotional images and a selective sleep benefit on emotional memory.

Additionally another of the central aspects to identifying the benefit of sleep on emotional memory is the involvement of REM sleep. For emotional memories it has been found that REM sleep is positively correlated with increased emotional memory performance (Nishida, Pearsall, Buckner, & Walker, 2009; Wagner, Kashyap, Diekelmann, & Born, 2007). REM sleep is thought to be important in the selective consolidation of emotional memories because it is considered to strengthen emotional memories and also attenuate the emotional tone associated with those memories (see Hutchison & Rathore, 2015 for a review; van der Helm & Walker, 2009). Again, like the other two experiments presented in chapters 2 and 4 we have no physiological measures of sleep to relate our findings to. Therefore it is entirely possible that sufficient REM sleep did not occur amongst our participants. This may of course indicate that in order to successfully conceptually replicate the selectivity of sleep-dependent memory consolidation on emotional memory two things are required, a benefit of emotion on

memory and sufficient REM sleep to drive the selective consolidation of those emotional memories.

At the beginning of this thesis I described two debates about the role of sleep in memory consolidation, the active account and the passive account. The main active account of sleep's role in memory consolidation proposed that sleep actively drives the stabilisation, strengthening and transformation of memories, and transfers them from the hippocampus to the Neocortex, otherwise known as the systems consolidation hypothesis (Born & Wilhelm, 2012; Diekelmann & Born, 2010; Marr 1971; McClelland, McNaughton, & O'Reilly 1995). The passive accounts of sleep's role in memory consolidation proposed that sleep is not actively involved in memory consolidation rather it only protects sleep from retroactive interference (Jenkins & Dallenbach, 1924). The pattern of evidence presented in this thesis cannot favour one explanation over the other. For example in two of the experiments we found greater recognition memory performance in our sleep condition compared to our wake condition. In this case we cannot determine whether sleep passively protected these memories or was actively involved in consolidating them since we do not have PSG measures of participants during sleep we cannot determine an active role. Additionally, in the emotional memory experiment we were capable of detecting an active role of sleep by examining whether sleep selectively consolidated emotional memories, but this was not successful. It is plausible that like the eyewitness identification experiment, the consolidative process that took place after wake were sufficient to maintain memory in these experiments; these do also occur with respect to emotional memory too (see Hutchinson & Rathore, 2015 for a review). Nevertheless in one case we did show that sleep may have an active role in memory consolidation and may drive pattern separation but this did not yield a benefit over and above a period of wake perhaps suggesting that by some other means wake too preserves memory, perhaps via a different mechanism, but to an equivalent extent. Importantly these unclear findings point to the

challenges faced by running online experiments alone without PSG when establishing whether sleep plays an active or passive role, but they do provide a reliable indicator of whether or not sleep benefits memory prior to bringing the experiment into the lab.

*Sleep norms.* Because there was so much data collected from the eyewitness memory experiment (Chapter 2), I could provide normative data for the Stanford Sleepiness Scale, Epworth Sleepiness Scale, and sleep duration. Although many differences were observed across gender, education, and age on the SSS, ESS and sleep duration, only one effect was meaningful: there was greater sleepiness in the evening compared to the morning. This identifies a potential crucial confound that maybe has gone undetected in research investigating the relationship between sleep and memory, as many experiments were not suitably powered to detect this effect.

### Critical Analysis of Findings

There are several possible reasons that may have prevented me from finding a sleep effect on recognition memory in this thesis (e.g., expected vs. unexpected retrieval, no control over sleep onset; both possibilities are discussed below; Nemeth, Gerbier & Janacsek, 2019). There are also reasons why the data presented in this thesis might be representative of the actual nature of the effect of sleep on recognition memory (e.g., time of day effects, sleepiness etc.). Nemeth et al. (2019) recently highlighted many of these issues. Each experimental chapter is discussed in turn against the possible reasons for the mixed findings presented in this thesis and for the purposes of theoretical exploration my findings are first discussed outside of potential confounds.

*Type of task.* One possible reason for the absence of an effect of sleep on memory is the type of task used. Researchers have argued and presented evidence that the benefits of sleep are more often observed in experiments where recall (e.g., Lahl, Wispel, Willigens &

Pietrowsky, 2008) or cued recall (e.g., Benson & Feinberg, 1977; Drosopoulos, Schulze, Fischer, & Born, 2007; Ellenbogen, Hulbert, Stickgold, Dinges & Thompson-Schill, 2006) tasks are used (e.g., Diekelmann, Wilhelm & Born, 2009). The pattern of results is more inconsistent for sleep experiments where a recognition memory task is used where some find an overall benefit of sleep relative to a period of wake (e.g., Koulack, 1997; Wagner, Kashyap, Diekelmann & Born, 2007) but the effect is small and others find it only under particular conditions (e.g., Drosopoulos, Wagner, & Born, 2005; Hu, Stylos-Allan & Walker, 2006).

Researchers have proposed that the less consistent benefits of sleep on recognition memory tasks reflect differences in the underlying impact of sleep on those tasks (Diekelmann, Wilhelm & Born, 2009), specifically the differences in neuroanatomical structures that the tasks use. For instance, Diekelmann, Wilhelm & Born (2009) proposed that recall tasks involve hippocampal activation, which is where newly encoded information is stored and subsequently reactivated during sleep. Comparatively the hippocampus is thought to be less related to processes involved in recognition memory tasks (Diekelmann, Wilhelm & Born, 2009). Therefore, the absence of a sleep effect in Chapter 2, in which the impact of sleep on eyewitness identifications was investigated, could simply be because the underlying processes occurring during sleep aid recall and not recognition.

However, this argument does not hold in light of the findings in Chapters 4 and 5 and other research examining the impact of sleep on recognition memory and the contributions of neuroanatomical processes on these tasks. In Chapters 4 and 5, sleep benefitted recognition memory overall in comparison to a period of wake. Therefore, it is unlikely that processes occurring during sleep do not benefit recognition memory performance. This is further corroborated by research indicating that recall and recognition do not involve separate neuroanatomical processes. Wixted and Squire (2004) found that in patients who have hippocampal damage, both recall and recognition memory are equally impaired. The argument



proposed by Diekelmann, Wilhelm & Born (2009) holds that only recall, not recognition, should be negatively affected. In fact, strong memories (e.g., recollection) which are more common in recall than recognition and weaker memories (e.g., familiarity) which are less common in recall are found to be equally involved in hippocampal functioning (e.g., Smith, Wixted & Squire, 2011; Wais, Wixted, Hopkins & Squire, 2006; Wais, Squire & Wixted, 2009). Thus, differences in the type of task may not be a suitable explanation for the inconsistencies in the data presented in this thesis.

*Expected vs. unexpected retrieval.* Another explanation is whether participants anticipate being tested on their memory. Wilhelm et al. (2011) asked participants to learn card locations and a sequential finger-tapping task. Participants either slept or remained awake during the day and were informed either that they were going to be tested or were not. When participants were informed that they will be tested performance was greater in both tasks in the sleep condition compared to the wake condition. A difference in performance on the tasks between the sleep and wake conditions did not emerge when participants were not informed that they would be tested. Wilhelm et al. (2011) and others (e.g., Fischer & Born, 2009) have proposed that retrieval anticipation facilitates and boosts the consolidation processes which occur during sleep. This may indicate that sleep does not enhance memory for all information that is studied and is instead selective based on what will be relevant for an individual's future. Therefore, one prediction might be that a benefit of sleep is only observed in comparison to a period of wake when participants are informed that their memory will be tested.

However, this account is not supported by the data presented in this thesis. For example, in Chapter 2, participants were informed that they would be tested on their memory for the perpetrator of the crime, and yet no difference in discriminability between sleep and wake conditions was found. In Chapter 4, the investigation of the impact of sleep on misinformation, old/new recognition memory was greater in the sleep conditions compared to the wake

conditions. In that experiment participants were informed that they would be tested on their memory. However, in Chapter 5, where the benefit of sleep on emotional vs. neutral stimuli was investigated, participants did not expect a memory test and participants in the sleep condition performed better on discriminability. If the benefits of sleep are conditional on retrieval anticipation then this effect would not occur, unless participants did anticipate being tested in this experiment, which is a possibility. Nonetheless whether or not participants expect to be tested on their memory does not have much support when taking the evidence presented in this thesis together.

*Retroactive interference occurring before sleep onset.* In all of the experiments the amount of time that occurred between participants finishing the task and their sleep onset was not controlled. This could be problematic as interference could occur, where new information encountered prior to sleeping may interfere with the information that was just encoded in the task at hand (Nemeth et al., 2019). However, one could argue that this would not diminish the benefits of sleep on memory because there should be less interference in the sleep condition. Nemeth et al. (2019) suggested that extended periods between learning and the onset of sleep could reduce the effect size that a researcher is able to detect. Some research does indicate that the benefits of sleep on memory are smaller when the time between learning and sleep onset is greater (Press, Casement, Pascual-Leone, & Robertson, 2005; Robertson, Press, & Pascual-Leone, 2005; Walker et al., 2003).

Could the lack of controlling the time between encoding and sleep explain the inconsistent findings of a sleep effect on memory throughout the results presented in this thesis? Consider the experiment presented in Chapter 2, where there was no benefit of sleep on discriminability. The fact that there was not a difference in discriminability between the sleep and wake group would unlikely be due to some interference occurring in the sleep group. The

experiment was powered to detect a very small effect. Nevertheless, this represents a particular challenge for conducting experiments online where control of this type is difficult.

*Threshold benefits of sleep on memory.* Another possible reason for the absence of a sleep effect in Chapter 2 could be related to the initial encoding strength. Recently Schoch et al. (2017) examined the impact of initial encoding memory strength on sleep and emotional memory. They proposed that the benefits of sleep might follow an inverted U function where the benefits of sleep are more likely to occur at medium levels of encoding strength. They cited two papers, which indicated that the benefits of memory may not occur when accuracy is very high or encoding strength is high (Creery, Oudiette, Anthony & Paller, 2015; Drosopoulos et al., 2007).

In Chapter 2, the participants had a 30-second exposure to the perpetrator's face, which is longer than is typical in eyewitness identifications where performance is not as high (e.g., 8-second exposure in Seale-Carlisle & Mickes, 2016). Comparatively, in the emotional memory experiment (Chapter 5), where memory performance was greater in the sleep compared to wake condition, exposure to each image was only 1-second per image. This may mean that encoding strength in the registered report is higher than encoding strength in the emotional memory experiment. If these differences resulted in an absence of an effect in one study and not the other, you would expect the magnitude of the differences in  $d'$  to differ between the studies. This is the case, the difference in  $d'$  between sleep and wake condition in Chapter 2 was 0.02, whereas the difference in  $d'$  between sleep and wake condition in Chapter 5 was 0.39. However, it is also important note that the findings in Chapter 5 may be confounded by time of day effects, so this difference may reflect the impact of time of day rather than initial encoding strength.

*Time of day effects and sleepiness.* There are two potential confounds present in some of the experiments in this thesis, time-of-day effects and sleepiness at study and test. These

factors were controlled for in case they could account for any differences detected between sleep and wake conditions. For example, a sleepiness confound may present itself in differences in sleepiness at study and test phases in a sleep experiment which finds a benefit of sleep on memory. In this scenario, sleepiness at study may be the same but sleepiness at test differs and participants in the wake condition were sleepier than participants in the sleep condition. Therefore, one explanation could be that the sleep group performed better because they were less sleepy at test compared to the wake group. A similar pattern could emerge for time-of day effects. For instance participants in the sleep group performed better because they were tested in the morning and participants in the wake group performed worse because they were tested in the evening.

In Chapter 3, sleepiness was greater in the evening compared to the morning, which could have impacted overall performance. The difference in sleepiness between morning and evening was a small to medium effect size. In order to detect an effect size ranging from .037 to .052 with 95% power, data from between 120-171 participants per condition would need to be collected. An effect larger than this was found in Chapter 4 with a smaller sample size. There were 60 participants in the sleep condition and 60 participants in the wake condition, and a large effect of  $d = 1.08$  was detected. Although smaller samples may inflate the size of the true effect (Button et al., 2013). Smaller samples may also not identify the effect that sleepiness differs depending on time of day. For example, with a much larger sample I detected differences between study and test in sleep and wake conditions (Chapter 2), the same was the case for the misinformation experiment (Chapter 4), but not for the emotional memory experiment, where the sample was smaller. This is useful information for researchers examining the relationship between sleep and memory, especially because many studies do not have sufficient sample sizes to rule out this effect.

Sleepiness did not appear to moderate the effect of sleep on memory in the eyewitness memory experiment (Chapter 2), where there was a difference in sleepiness depending on the time of day. Additionally, although a difference in sleepiness was found between study and test in the misinformation experiment (Chapter 4), it was in the conditions where no significant difference between sleep and wake was found. No differences in study and test in sleepiness between sleep and wake conditions was found, but it remains a possibility that sleepiness could impact some of the findings in this study, but power was too low to detect significant differences.

In sleep experiments, it is important to control for time of day confounds, which was done in the experiments in this thesis by collecting data from AM and PM controls in a PM AM design (Nemeth et al., 2019). By including these controls, two important findings were revealed. In the emotional memory experiment (Chapter 5), a non-significant *p*-value but medium-sized effect was revealed between AM and PM control conditions on recognition memory. Performance was greater in the AM control condition compared to the PM control condition. This finding limits the conclusions that can be made on differences in *d'* between the sleep and wake conditions. However, the same reasoning that applies to that of assessing sleepiness with smaller sample sizes applies to potential time of day confounds, where the effect could be overestimated. Indeed, in the larger sample sizes in other experiments effects of the same magnitude were not detected. However, this possibility should not be discounted given that previous research has indicated that time of day effects can occur with samples larger than my own (e.g., Rothen & Meier, 2017; *N* = 115-118 per condition).

The fact that many studies do not include time of day controls and are not suitably powered to detect differences in sleepiness at study and test is problematic for interpreting results. It is entirely plausible that the effects may not be real effects of sleep on memory, but confounds of sleepiness and time-of-day. These factors should always be taken into

consideration in future experiments so that interpretations of the results are clearly of a sleep benefit on memory and not due to sleepiness or time of day confounds (e.g., Nemeth et al., 2019).

### Limitations

*Specification criteria.* In all of the experiments, participants were asked about their physical (e.g., neurological disorders) and mental health (e.g., psychiatric disorders) and various other factors, which may have affected their memory. Such factors may have also affected an ability to detect a benefit of sleep on memory compared to a period of wake (e.g., napping, sleep duration < 6 hours, etc.). After controlling for these factors, there was not a benefit of sleep on discriminability in the eyewitness experiment (Chapter 2), but there was a benefit of sleep on old/new recognition memory in the misinformation experiment (Chapter 4). There was also a benefit of sleep on recognition memory in the emotional memory experiment (Chapter 5) despite including the data from those individuals with factors that could impact memory (those who were excluded in the other experiments). However, by including these individuals I may have also inhibited my ability to detect a benefit of sleep on emotional information over neutral information. Perhaps sleep consolidation in those with sleep disorders and other factors does not yield as strong an effect and weakened the overall effect of sleep on emotional memory in this sample. However, it is unclear how the benefits of sleep on old/new recognition memory would be found overall but not for emotional stimuli by including these individuals.

Following precedence (e.g., Tamminen, Ralph & Lewis, 2013), in all of the experiments data from participants who may have factors that impact their memory were excluded. We do not, however, know whether excluding or including participants on the basis of these criteria impacts our ability to determine the effect of sleep on recognition memory. Indeed, not all researchers employ the same exclusionary criteria and still detect the benefits of sleep on

memory. Many exclusionary criteria are defensible based on prior findings, but should not be based on intuition.

Exclusionary criteria are only one source of experimental specifications, others also exist, including the type of analyses (e.g., within- or between-subjects analyses), the type of task used (e.g., recognition, free recall), and the specifications within a task (e.g., exposure duration, retention interval etc.) (Simonsohn, Simons & Nelson, 2015). Specification variations may be a source of inconsistency of results in published experiments examining the impact of sleep on recognition memory and may also explain the inconsistency in of the findings presented in the current thesis.

An important challenge is to determine whether inconsistencies in the research reported in this thesis and other published sleep research have arisen out of arbitrary experimenter decisions (e.g., selectively reporting of data or stopping data collection early) or decisions that underpin the theories being tested (Simonsohn, Simons & Nelson, 2015). One way of doing this is by conducting specification curve analysis (Simonsohn, Simons & Nelson, 2015). There are three steps to specification curve analysis. The first is to select the specifications of interest that researchers may use within their experiment. The second step is to randomly combine these specifications and test which ones yield an increased benefit of sleep on memory and which of those combinations yield a significant difference. The third step is to determine whether the results of those specifications are inconsistent with the null. Unfortunately, I am not able to conduct this analysis on the data presented in this thesis.

Specifically, I would be interested in iterating between random combinations of my exclusionary criteria to see if any of them are required to detect a benefit of sleep on memory. This cannot be achieved with data collected in the eyewitness experiment (Chapter 2) because those who failed to meet the criteria did not participate. Nor can this be achieved in the misinformation experiment (Chapter 4) or the emotional memory experiment (Chapter 5)

because there are not enough individuals who did not meet those criteria within those datasets, so the analyses would be underpowered. This limits the ability to determine one potential source of the inconsistencies found in my experiments, and is a source of future investigation for researchers investigating the relationship between sleep and memory.

*Measurement of sleep physiology.* The experiments in this thesis were conducted online, instead of in the lab to achieve greater power. I therefore did not collect measurements of sleep physiology (e.g., slow wave sleep, sleep spindles etc.). This further limits the conclusions I can make about the involvement of sleep in recognition memory performance, particularly where those effects are found. For example, in Chapter 5, I found that sleep benefits recognition memory performance relative to an equivalent period of wake. The question here, time of day effects aside, is whether sleep plays a passive role or active role in preserving memory. One way of determining this is to measure the relationship between sleep physiology and memory performance. This is typically achieved by correlating memory performance with slow wave sleep or spindle activity (e.g., Hanert et al., 2017; Pardilla-Delgado & Payne, 2017; Tamaki, Matsuoka, Nittono, & Hori, 2008). If a relationship is found between memory and sleep physiology, then one conclusion might be that sleep plays an active role in memory consolidation (see Walker, 2009). However, if sleep does not, then the role of sleep may be passive, where sleep protects memory from interference, but does not actively drive consolidation of those memories (See Ellenbogen, Payne & Stickgold, 2006). Such measurements have informed theories about sleep's role in memory consolidation (e.g., Diekelmann & Born, 2010; Stickgold, 2005), and if available could enable me to elaborate further on some of the conclusions drawn in this thesis.

However, it is important to acknowledge that collection of sleep physiology and correlating it with memory performance has been heavily criticised as a method to understand the active involvement of sleep in memory consolidation (e.g., Ackermann et al., 2015;



Mantua, 2018; Nemeth et al., 2019; Pan & Rickard, 2015). The major source of criticism of these analyses comes from a lack of power. Often, researchers conduct sleep and memory experiments in the lab using small sample sizes. One way to assess the impact of lack of power and how that can lead to overestimations of the effect of sleep stages or activity on memory is to conduct simulations.

The simulations were conducted on data from Clemens, Fabo and Halasz (2006). Clemens et al. recorded spindle activity during sleep in 15 individuals and found a positive correlation between visuospatial memory and spindle activity. To estimate the data, 15 data points for memory performance were drawn from a normal distribution with  $M = 0.00$  and a  $SD = 2.80$  and 15 data points for spindle activity were drawn from a normal distribution with a  $M = 1623.00$  and a  $SD = 569.00$ . Figure 10A shows the results of one simulation and effectively demonstrates that if the experiment were conducted again, the significant positive correlation between memory retention and spindle activity may replicate. Running the simulation again shows that the correlation is no longer significant and is, in fact, trending in the opposite direction, as shown in Figure 10B.

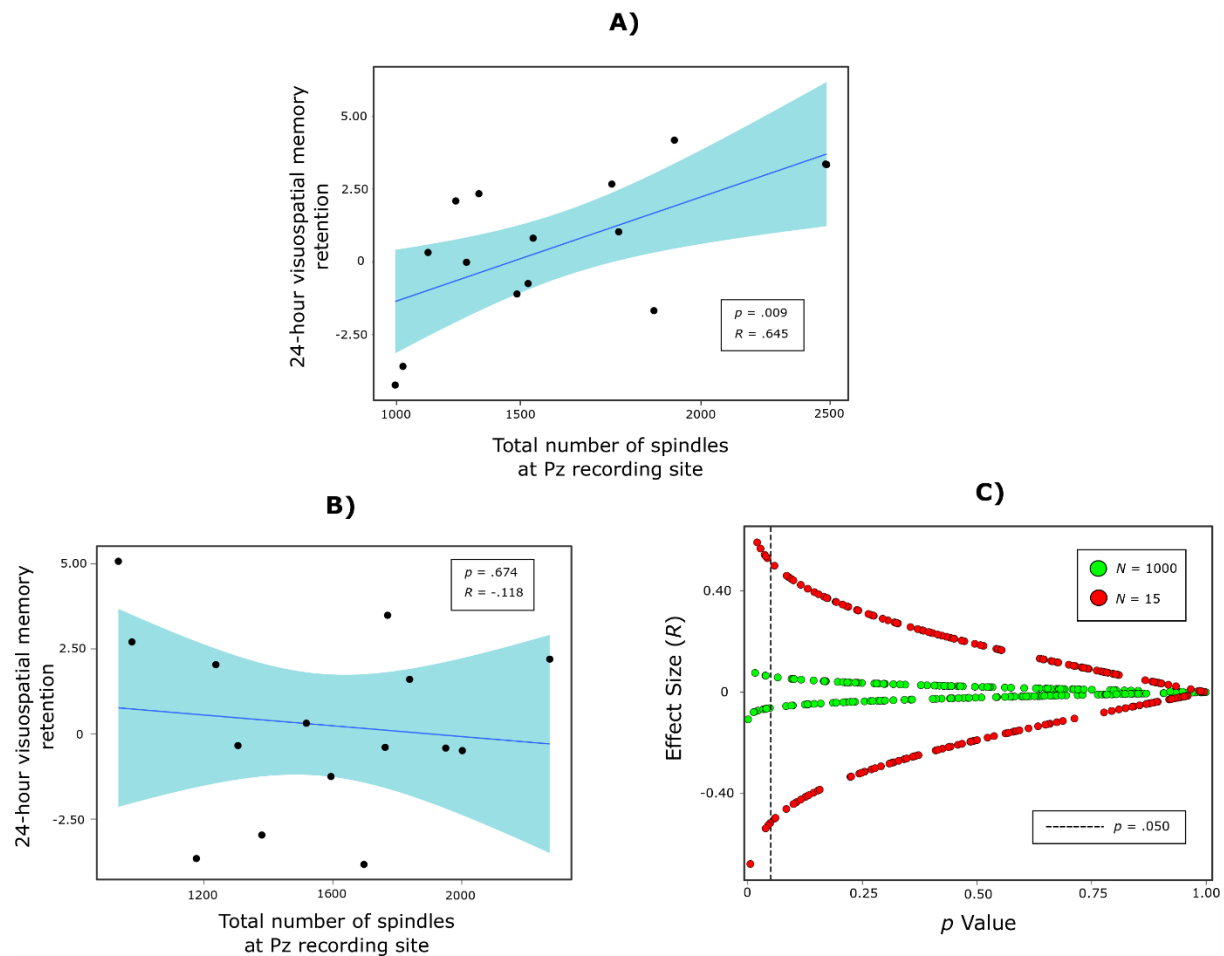


Figure 10. A) One simulation based on Clemens's et al. (2006) data, error bars represent SEM. B) Another simulation based on Clemens's et al. (2006) data, error bars represent SEM. C) Effect sizes and corresponding p values across 200 simulations of Clemens's et al. (2006) data with a small ( $N = 15$ ) and large sample size ( $N = 1000$ ).

With respect to Clemens et al. (2006), it may be the case that the latter is more typical for two reasons, depicted in Figure 10C. Figure 10C shows the p-value and corresponding effect size for each simulation conducted using the data reported in Clemens et al. The dashed line represents the significance criterion,  $p < .050$ . When the data were simulated using a small sample size the effect size was larger than when the data were simulated using a much larger sample size. Additionally, when using a small sample, it is easy to appreciate that this experiment would only replicate a small number of times, as only a small number of

simulations are significant in the same direction. Most simulations are not significant, and some are in the opposite direction. This is because small samples increase the chance of making a type I error when the null hypothesis, which is actually true, is rejected. It should also be pointed out that the same is true when the data are simulated with a larger sample, but this may be because the means and standard deviations reported in Clemens et al. reflect a sampling error. Collecting new data with a larger sample size may yield fewer false positives because more reliable means and standard deviations can be achieved. Regardless, it is clear that correlations based on small samples are likely to yield false positive results and overestimate effect sizes.

It is well known that smaller sample sizes can overestimate the effect that a researcher intends to detect (e.g., Button et al., 2013). For example, in support of the differing simulations in Figure 10C, Ackerman et al. (2015) demonstrated that when larger sample sizes are used ( $N = 885$ ) the relationship between sleep stages and memory performance disappears. In fact, it is possible that many of the correlations reported in the literature examining the relationship between sleep and memory reflect false positives because small samples are used. In response to this Nemeth et al. (2019) have recommended when performing such correlations, they should be corrected for multiple comparisons and that sleep physiology data should be shared on the Open Science Framework. This would enable researchers to compile and aggregate data across studies to determine the effect sizes related to sleep stages, activity, and memory performance. I recommend that a further step should be taken, that is a large-scale collaboration of researchers to investigate the relationship between sleep and memory, both behaviourally and physiologically.

### Future Directions

*The Many Zs Initiative.* Throughout this thesis, I have explored potential reasons for the absence of strong evidence for the impact of sleep on recognition memory (see Table 1).

Two themes were identified: a lack of replication and questionable power. These are problematic because our theoretical understanding of the impact of sleep is based on experiments that have not been re-examined. This is especially important when small sample sizes are used. These experiments should be replicated because smaller samples can lead to false positives and overestimation of effect sizes.

It is plausible that this is the case in the sleep literature as most studies report some form of benefit of sleep on memory, and very few report no effect of sleep compared to wake (e.g., Rickard, Cai & Jones, 2008). Many researchers may be finding it difficult to demonstrate effects that are thought to be well-established, and are wasting resources (e.g., money, time, etc.) in their attempts. Alternatively, they could be finding no effects but have difficulty publishing those results. Such resources may be put to better use by performing a large-scale replication attempt. It could be the case that these previously reported effects replicate, but in the context of the current publication bias, this may be unlikely. Therefore, there is an argument in favour of conducting a large-scale replication attempt of previous research.

I therefore propose a large-scale replication effort named “The Many Zs Initiative,” modelled after recent large-scale replication collaborations (e.g. The Many Labs 2, 2018; The Rahal & Open Science Collaboration, 2015). This initiative would use first the online approach used in this thesis and then include lab-based work after behavioural effects were replicated. First researchers in the sleep and memory community would be contacted to determine their interest in performing a large-scale replication effort into the effects of sleep on memory. Once researchers have been gathered, potential targeted experiments to be replicated would be selected and polled by sleep researchers to determine which ones researchers would be most interested in seeing replicated. Once chosen, the hypotheses, methods, analyses and analysis scripts would be created and pre-registered with the Open Science Framework, or a journal in the form of a registered report.

The chosen experiments would not be brought into the lab at first, instead large-scale online experiments would be conducted, like the registered report presented in Chapter 2. By collecting large behavioural datasets when examining the impact of sleep on memory enables researchers to identify a more reliable estimate of the true effect of the behavioural manipulation. Once an effect size is established the experiment can then be brought into the lab and powered accordingly, which saves time, money and resources in the end. If no effect of sleep on memory is found then that experiment will not be brought into the lab. The in lab experiments would be conducted to not only attempt to replicate the online finding but to also determine the relationship between sleep physiology and memory performance. As the program for the experiment would be coded online, it could be easily distributed and replicated across labs. Therefore, using modern technologies (e.g., Amazon Web Services, mTurk) we can move past running underpowered sleep experiments and towards running large-scale reliable examinations of the impact of sleep on memory.

My preference of experiments to replicate would be those examining the impact of sleep on recognition memory. This is because there appears to be the greatest inconsistency in those results. Sometimes sleep benefits weak memories (e.g., Hu et al., 2008) or aids strong memories (e.g., Schoch et al., 2017; Schönauer et al., 2013). This is a particularly interesting line of replication as it would enable us to determine whether published findings are replicable and enable us to have greater faith in or update our theoretical understanding of the impact of sleep on memory. One line of future investigation is the impact of sleep on memory strength. Many studies in sleep research have manipulated encoding memory strength (e.g., emotional vs. neutral, reward vs. no reward) and fewer have examined retrieval memory strength (e.g., recollection vs familiarity). One feature, which is missed from many of these studies, is equating for confidence. Determining the impact of sleep on reward or emotion or recollection vs familiarity suggests that the impact of sleep on memory is an all or none process, either

memories benefit from consolidation or they do not. One way around this issue is by equating for confidence. It may be the case that differences in memory for high or low reward are not present at high levels of confidence, but perhaps are at lower levels. Nevertheless, this theoretical advancement relies on the replication of studies that identified these effects originally.

## Conclusion

The experiments presented in this thesis do not provide consistent support for the benefits of sleep on recognition memory, specifically eyewitness memory and emotional memory. The first experiment, in Chapter 2, showed no evidence of a benefit of sleep on memory, whereas the experiments presented in Chapters 4 and 5 do show a benefit of sleep on recognition memory. In these experiments it was difficult to determine whether sleep has an active or passive role in the memory advantages of sleep found here since we do not have physiological measures and our manipulations which could reveal an active role of sleep failed. Therefore whether these findings have implications for theory on either side of the debate is unclear. Moreover, the experiment in Chapter 4 was conducted using a problematic design where despite there being a 12-hour difference in retention intervals the  $d'$  values were not different and the experiment in Chapter 5 may be limited due to a potential time of day effect. That aside, I explored a number of alternative reasons for the results, however, none account for the inconsistencies in the data presented in this thesis. But, it should be acknowledged that my findings only pertain to the 12-hour AM:PM-PM:AM design and do not necessarily apply to other designs such as napping studies where the conditions in which the benefits of sleep are observed differ greatly. For example, napping studies eliminate time-of-day effects because study and testing for napping and no napping groups can occur at the same time. It may also eliminate the sleepiness confounds which could be argued to be caused by time-of-day effects. Therefore, running the same experiments presented here, but in a napping

paradigm, may remove some of the doubt I cast on my some of my findings because of sleepiness and time-of-day-confounds. Although, the fact that the benefits of sleep on memory emerged in the presence of possible confounds, may indicate that many of the findings (including the confounds) presented in this thesis need to be re-examined. Nevertheless, it is my view that to advance our theoretical understanding of the impact of sleep on memory large-scale replication efforts must be pursued to reliably determine and demonstrate the impact that sleep may have on memory consolidation.

## Appendices

### Appendix A

Calculation of  $d_a$  varied depending on the type of memory being examined:

#### *Old/new recognition memory*

$d_a$  was calculated to determine participants' ability to distinguish between targets (i.e. events that occurred in the original event) and lures (i.e. events that did not occur)

$$d_a = \frac{(\mu_{\text{Targets}} - \mu_{\text{Lures}})}{\sqrt{\frac{\sigma_{\text{Targets}}^2 + \sigma_{\text{Lures}}^2}{2}}}$$

where  $\mu_{\text{Targets}}$  referred to the proportion of “yes” responses made to targets (i.e. a hit) and  $\mu_{\text{Lures}}$  referred to the proportion of “yes” responses made to lures (i.e. a false alarm).  $\sigma_{\text{Targets}}^2$  and  $\sigma_{\text{Lures}}^2$  referred to the standard deviations of the proportion of “yes” responses made to targets and lures.

#### *Memory for misinformation*

$d_a$  was calculated to determine participants' ability to distinguish between Misinformation (i.e. events that occurred in the misinformation) and lures (i.e. events that did not occur).

$$d_a = \frac{(\mu_{\text{Misinformation}} - \mu_{\text{Lures}})}{\sqrt{\frac{\sigma_{\text{Misinformation}}^2 + \sigma_{\text{Lures}}^2}{2}}}$$

where  $\mu_{\text{Misinformation}}$  referred to the proportion of “yes” responses made to misinformation (i.e. a hit) and  $\mu_{\text{Lures}}$  referred to the proportion of “yes” responses made to lures (i.e. a false



alarm).  $\sigma_{\text{Misinformation}}^2$  and  $\sigma_{\text{Lures}}^2$  referred to the standard deviations of the proportion of “yes” responses made to misinformation and lures.

### *Source memory*

$d_a$  was calculated to determine participants’ ability to distinguish between information that occurred in the study phase (i.e. pictures) and information that occurred in the misinformation phase (i.e. sentences)

$$d_a = \frac{(\mu_{\text{Pictures}} - \mu_{\text{Sentences}})}{\sqrt{\frac{\sigma_{\text{Pictures}}^2 + \sigma_{\text{Sentences}}^2}{2}}}$$

where  $\mu_{\text{Pictures}}$  referred to the proportion of “picture” responses made to targets (i.e. information that occurred in the study phases) and  $\mu_{\text{Sentences}}$  referred to the proportion of “picture” responses made to misinformation (i.e. information that occurred in the misinformation phase).  $\sigma_{\text{Pictures}}^2$  and  $\sigma_{\text{Sentences}}^2$  referred to the standard deviations of the proportion of “yes” responses made to targets and misinformation.

## References

- Ackermann, S., Hartmann, F., Papassotiropoulos, A., de Quervain, D. J., & Rasch, B. (2015). No associations between interindividual differences in sleep parameters and episodic memory consolidation. *Sleep*, 38, 951-959.
- Adan, A., & Almirall, H. (1991). Horne & Östberg morningness-eveningness questionnaire: A reduced scale. *Personality and Individual Differences*, 12, 241-253.
- Ahn, H., Wijaya, M. E., & Esmero, B. C. (2014). A systemic smartphone usage pattern analysis: focusing on smartphone addiction issue. *International Journal of Multimedia Ubiquitous Engineering*, 9, 9-14.
- Allanson, F., & Ecker, U. K. (2017). No evidence for a role of reconsolidation in updating of paired associates. *Journal of Cognitive Psychology*, 29, 912-919.
- Alóe, F., Pedroso, A., & Tavares, S. M. (1997). Epworth Sleepiness Scale outcome in 616 Brazilian medical students. *Arquivos de Neuro-psiquiatria*, 55, 220-226.
- Altevogt, B. M., & Colten, H. R. (Eds.). (2006). *Sleep disorders and sleep deprivation: an unmet public health problem*. Washington, DC: National Academies Press.
- Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, 410, 366-369.
- Arechar, A. A., Gächter, S., & Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics*, 21, 99-131.
- Atienza, M., & Cantero, J. L. (2008). Modulatory effects of emotion and sleep on recollection and familiarity. *Journal of Sleep Research*, 17, 285-294.

- 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 & Memory*, 89, 76-80.
- Baglioni, C., Battagliese, G., Feige, B., Spiegelhalder, K., Nissen, C., Voderholzer, U., ... & Riemann, D. (2011). Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. *Journal of Affective Disorders*, 135, 10-19.
- BaHammam, A. S., Almestehi, W., Albatli, A., & AlShaya, S. (2011). Distribution of chronotypes in a large sample of young adult Saudis. *Annals of Saudi Medicine*, 31, 183-186.
- Bailes, C., Caldwell, M., Wamsley, E. J., & Tucker, M. A. (2020). Does sleep protect memories against interference? A failure to replicate. *PloS one*, 15, e0220419.
- Baker, L. D., Baldwin, D. S., & Garner, M. (2015). Daytime intrusive thoughts and subjective insomnia symptoms. *Psychiatry Research*, 229, 1038-1042.
- Banks, S., Barnes, M., Tarquinio, N., Pierce, R. J., Lack, L. C., & Doug McEvoy, R. (2004). Factors associated with maintenance of wakefulness test mean sleep latency in patients with mild to moderate obstructive sleep apnoea and normal subjects. *Journal of Sleep Research*, 13, 71-78.
- Baran, B., Pace-Schott, E. F., Ericson, C., & Spencer, R. M. (2012). Processing of emotional reactivity and emotional memory over sleep. *Journal of Neuroscience*, 32, 1035-1042.
- Barrett, T. R., & Ekstrand, B. R. (1972). Effect of sleep on memory: III. Controlling for time-of-day effects. *Journal of Experimental Psychology*, 96, 321-327.

- Bartlett, J. C. (1977). Effects of immediate testing on delayed retrieval: Search and recovery operations with four types of cue. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 719-732.
- Bartlett, J. C., & Tulving, E. (1974). Effects of temporal and semantic encoding in immediate recall upon subsequent retrieval. *Journal of Verbal Learning & Verbal Behavior*, 13, 297-309.
- Bartneck, C., Duenser, A., Moltchanova, E., & Zawieska, K. (2015). Comparing the similarity of responses received from studies in Amazon's Mechanical Turk to studies conducted online and with direct recruitment. *PloS One*, 10, e0121595.
- Bastos, M., & Mercea, D. (2018). The public accountability of social platforms: lessons from a study on bots and trolls in the Brexit campaign. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376.
- Bauters, F., Rietzschel, E. R., Hertegonne, K. B., & Chirinos, J. A. (2016). The link between obstructive sleep apnea and cardiovascular disease. *Current Atherosclerosis Reports*, 18, 1.
- Beaudreau, S. A., Spira, A. P., Stewart, A., Kezirian, E. J., Lui, L. Y., Ensrud, K., ... & Study of Osteoporotic Fractures. (2012). Validation of the Pittsburgh Sleep Quality Index and the Epworth Sleepiness Scale in older black and white women. *Sleep Medicine*, 13, 36-42.
- Benjamin, A. S., Diaz, M., & Wee, S. (2009). Signal detection with criterion noise: applications to recognition memory. *Psychological Review*, 116, 84-115.
- Benson, K., & Feinberg, I. (1977). The beneficial effect of sleep in an extended Jenkins and Dallenbach paradigm. *Psychophysiology*, 14, 375-384.

Berres, S. & Erdfelder, E. 2018 Welche Rolle Spielt schlaf im episodischen gedachtnis?

Vorläufige ergebnisse einer meta-analyse. Poster presented at the annual Deutsche Gesellschaft für psychologie in Frankfurt, Germany.

Bewick, V., Cheek, L., & Ball, J. (2003). Statistics review 8: Qualitative data—tests of association. *Critical Care*, 8, 46.

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

Bloch, K. E., Schoch, O. D., Zhang, J. N., & Russi, E. W. (1999). German version of the Epworth sleepiness scale. *Respiration*, 66, 440-447.

Born, J. (2010). Slow-wave sleep and the consolidation of long-term memory. *The World Journal of Biological Psychiatry*, 11, 16-21.

Born, J., & Wilhelm, I. (2012). System consolidation of memory during sleep. *Psychological Research*, 76, 192-203.

Borota, D., Murray, E., Keceli, G., Chang, A., Watabe, J. M., Ly, M., ... & Yassa, M. A. (2014). Post-study caffeine administration enhances memory consolidation in humans. *Nature neuroscience*, 17, 201-203.

Brawn, T. P., Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2010). Consolidating the effects of waking and sleep on motor-sequence learning. *Journal of Neuroscience*, 30, 13977-13982.

Bridge, D. J., & Paller, K. A. (2012). Neural correlates of reactivation and retrieval-induced distortion. *Journal of Neuroscience*, 32, 12144-12151.

- Burnham, W. H. (1903). Retroactive amnesia: illustrative cases and a tentative explanation. *The American Journal of Psychology*, 14, 382-396.
- Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14, 365-376.
- Caci, H., Deschaux, O., Adan, A., & Natale, V. (2009). Comparing three morningness scales: age and sex effects, structure and cut-off criteria. *Sleep Medicine*, 10, 240-245.
- Cai, D. J., & Rickard, T. C. (2009). Reconsidering the role of sleep for motor memory. *Behavioral Neuroscience*, 123, 1153.
- Cairney, S. A., Lindsay, S., Paller, K. A., & Gaskell, M. G. (2018). Sleep preserves original and distorted memory traces. *Cortex*, 99, 39-44.
- Calvillo, D. P., Parong, J. A., Peralta, B., Ocampo, D., & Van Gundy, R. (2016). Sleep increases susceptibility to the misinformation effect. *Applied Cognitive Psychology*, 30, 1061-1067.
- Cappuccio, F. P., D'Elia, L., Strazzullo, P., & Miller, M. A. (2010). Quantity and quality of sleep and incidence of type 2 diabetes: A systematic review and meta-analysis. *Diabetes Care*, 33, 414-420.
- Carciofo, R., Du, F., Song, N., Qi, Y., & Zhang, K. (2012). Age-related chronotype differences in Chinese, and reliability assessment of a reduced version of the Chinese Morningness Eveningness Questionnaire. *Sleep and Biological Rhythms*, 10, 310-318.
- Carlson, C. A., & Carlson, M. A. (2014). An evaluation of lineup presentation, weapon presence, and a distinctive feature using ROC analysis. *Journal of Applied Research in Memory & Cognition*, 3, 45-53.

- Carlson, C. A., Dias, J. L., Weatherford, D. R., & Carlson, M. A. (2017). An investigation of the weapon focus effect and the confidence–accuracy relationship for eyewitness identification. *Journal of Applied Research in Memory and Cognition*, 6, 82-92.
- 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.
- Center for Disease Control and Prevention. (2005). Percentage of adults who reported an average of  $\leq 6$  hours of sleep per 24-hour period, by sex and age group—United States 1985 and 2004. *Morbidity and Mortality Weekly Report*, 54, 933.
- Centers for Disease Control and Prevention. (2008). QuickStats: percentage of adults aged  $> 18$  years who reported an average of  $< 6$  hours of sleep per 24-hour period, by sex and age group—National Health Interview Survey, United States, 1985 and 2006. *Morbidity and Mortality Weekly Report*, 57, 209.
- Centers for Disease Control and Prevention. (2009). Perceived insufficient rest or sleep among adults—United States, 2008. *Morbidity and Mortality Weekly Report*, 58, 1175-1179.
- Chambers, C. D. (2013). Registered reports: a new publishing initiative at Cortex. *Cortex*, 49, 609-610.
- Chelminski, I., Petros, T. V., Plaud, J. J., & Ferraro, F. R. (2000). Psychometric properties of the reduced Horne and Ostberg questionnaire. *Personality and Individual Differences*, 29, 469-478.

- Chung, S., Bohnen, N. I., Albin, R. L., Frey, K. A., Müller, M. L., & Chervin, R. D. (2013). Insomnia and sleepiness in Parkinson disease: Associations with symptoms and comorbidities. *Journal of Clinical Sleep Medicine*, 9, 1131-1137.
- Cipolli, C., Mazzetti, M., & Plazzi, G. (2013). Sleep-dependent memory consolidation in patients with sleep disorders. *Sleep Medicine Reviews*, 17, 91-103.
- Clark, S. E. (2012). Costs and benefits of eyewitness identification reform: Psychological science and public policy. *Perspectives on Psychological Science*, 7, 238-259.
- Clark, S. E., Brower, G. L., Rosenthal, R., Hicks, J. M., & Moreland, M. B. (2013). Lineup administrator influences on eyewitness identification and eyewitness confidence. *Journal of Applied Research in Memory & Cognition*, 2, 158-165.
- Clark, S. E., Marshall, T. E., & Rosenthal, R. (2009). Lineup administrator influences on eyewitness identification decisions. *Journal of Experimental Psychology: Applied*, 15, 63.
- Clemens, Z., Fabo, D., & Halasz, P. (2005). Overnight verbal memory retention correlates with the number of sleep spindles. *Neuroscience*, 132, 529-535.
- Clemens, Z., Fabó, D., & Halász, P. (2006). Twenty-four hours retention of visuospatial memory correlates with the number of parietal sleep spindles. *Neuroscience Letters*, 403, 52-56.
- Colloff, M. F., Wade, K. A., & Strange, D. (2016). Unfair lineups make witnesses more likely to confuse innocent and guilty suspects. *Psychological Science*, 27, 1227-1239.
- Cox, R. C., & Olatunji, B. O. (2016). A systematic review of sleep disturbance in anxiety and related disorders. *Journal of Anxiety Disorders*, 37, 104-129.



- Cox, R., Van Bronkhorst, M. L., Bayda, M., Gomillion, H., Cho, E., Parr, M. E., ... & Stickgold, R. (2018). Sleep selectively stabilizes contextual aspects of negative memories. *Scientific Reports*, 8.
- Creery, J. D., Oudiette, D., Antony, J. W., & Paller, K. A. (2015). Targeted memory reactivation during sleep depends on prior learning. *Sleep*, 38, 755-763.
- Danielsson, K., Markström, A., Broman, J. E., von Knorring, L., & Jansson-Fröjmark, M. (2016). Delayed sleep phase disorder in a Swedish cohort of adolescents and young adults: prevalence and associated factors. *Chronobiology International*, 33, 1331-1339.
- Daurat, A., Terrier, P., Foret, J., & Tiberge, M. (2007). Slow wave sleep and recollection in recognition memory. *Consciousness and Cognition*, 16, 445-455.
- Davidson, T. J., Kloosterman, F., & Wilson, M. A. (2009). Hippocampal replay of extended experience. *Neuron*, 63, 497-507.
- De Cock, V. C., Bayard, S., Jaussent, I., Charif, M., Grini, M., Langenier, M. C., ... & Dauvilliers, Y. (2014). Daytime sleepiness in Parkinson's disease: A reappraisal. *PLoS One*, 9, e107278.
- Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, 11, 114.
- 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, 381.
- Diekelmann, S., Wilhelm, I., & Born, J. (2009). The whats and whens of sleep-dependent memory consolidation. *Sleep Medicine Reviews*, 13, 309-321.

- Dobolyi, D. G., & Dodson, C. S. (2013). Eyewitness confidence in simultaneous and sequential lineups: A criterion shift account for sequential mistaken identification overconfidence. *Journal of Experimental Psychology: Applied*, 19, 345.
- Dodson, C. S., & Dobolyi, D. G. (2016). Confidence and Eyewitness Identifications: The Cross-Race Effect, Decision Time and Accuracy. *Applied Cognitive Psychology*, 30, 113-125.
- Donaldson, W. (1971). Output effects in multitrial free recall. *Journal of Verbal Learning & Verbal Behavior*, 10, 577-585.
- 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, 169-183.
- Drosopoulos, S., Wagner, U., & Born, J. (2005). Sleep enhances explicit recollection in recognition memory. *Learning & Memory*, 12, 44-51.
- Drosopoulos, S., Windau, E., Wagner, U., & Born, J. (2007). Sleep enforces the temporal order in memory. *PLoS One*, 2, e376.
- DSM-V-TR, A. P. A. (2013). *Diagnostic and statistical manual of mental disorders*. Arlington, VA: American Psychiatric Association.
- Dudai, Y., Karni, A., & Born, J. (2015). The consolidation and transformation of memory. *Neuron*, 88, 20-32.
- Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, 18, 35-39.

- Dunlap, R. E., McCright, A. M., & Yarosh, J. H. (2016). The political divide on climate change: Partisan polarization widens in the US. *Environment: Science and Policy for Sustainable Development*, 58, 4-23.
- Dunn, J. C. (2004). Remember-know: a matter of confidence. *Psychological Review*, 111, 524-542.
- Dunn, J. C. (2008). The dimensionality of the remember-know task: a state-trace analysis. *Psychological Review*, 115, 426-446.
- Ebbinghaus, H. (1885). *Über das Gedchtnis. Untersuchungen zur experimentellen Psychologie*. Leipzig: Duncker & Humblot; the English edition is Ebbinghaus, H. (1913). *Memory. A Contribution to Experimental Psychology*. New York: Teachers College, Columbia University
- Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology*. New York: Teachers College, Columbia University.
- Ecker, U. K., Brown, G. D., & Lewandowsky, S. (2015). Memory without consolidation: Temporal distinctiveness explains retroactive interference. *Cognitive Science*, 39, 1570-1593.
- 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, 1290-1294.
- 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, 1290-1294.

- 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, 716-722.
- Ellis, B. W., Johns, M. W., Lancaster, R., Raptopoulos, P., Angelopoulos, N., & Priest, R. G. (1981). The St. Mary's Hospital sleep questionnaire: a study of reliability. *Sleep*, 4, 93-97.
- Fatani, A., Al-Rouqi, K., Al Towairky, J., Ahmed, A. E., Al-Jahdali, S., Ali, Y., ... & Al-Jahdali, H. (2015). Effect of age and sex in the prevalence of excessive daytime sleepiness among a sample of the Saudi population. *Journal of Epidemiology and Global Health*, 5, 59-66.
- Feld, G. B., & Born, J. (2017). Sculpting memory during sleep: Concurrent consolidation and forgetting. *Current Opinion in Neurobiology*, 44, 20-27.
- Feld, G. B., & Diekelmann, S. (2015). Sleep smart—optimizing sleep for declarative learning and memory. *Frontiers in Psychology*, 6, 622.
- Feld, G. B., Besedovsky, L., Kaida, K., Münte, T. F., & Born, J. (2014). Dopamine D2-like receptor activation wipes out preferential consolidation of high over low reward memories during human sleep. *Journal of Cognitive Neuroscience*, 26, 2310-2320.
- Fenn, K. M., & Hambrick, D. Z. (2012). Individual differences in working memory capacity predict sleep-dependent memory consolidation. *Journal of Experimental Psychology: General*, 141, 404-410.
- Fenn, K. M., Gallo, D. A., Margoliash, D., Roediger, H. L., & Nusbaum, H. C. (2009). Reduced false memory after sleep. *Learning & Memory*, 16, 509-513.

- Fernández-Ruiz, A., Oliva, A., de Oliveira, E. F., Rocha-Almeida, F., Tingley, D., & Buzsáki, G. (2019). Long-duration hippocampal sharp wave ripples improve memory. *Science*, 364, 1082-1086.
- Fischer, S., & Born, J. (2009). Anticipated reward enhances offline learning during sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1586-1593.
- Fischer, S., Nitschke, M. F., Melchert, U. H., Erdmann, C., & Born, J. (2005). Motor memory consolidation in sleep shapes more effective neuronal representations. *Journal of Neuroscience*, 25, 11248-11255.
- Ford, E. S., Wheaton, A. G., Cunningham, T. J., Giles, W. H., Chapman, D. P., & Croft, J. B. (2014). Trends in outpatient visits for insomnia, sleep apnea, and prescriptions for sleep medications among US adults: Findings from the National Ambulatory Medical Care Survey 1999-2010. *SLEEP* 37, 1283-1293.
- Ford, J. B. (2017). Amazon's Mechanical Turk: a comment. *Journal of Advertising*, 46, 156-158.
- Forstmeier, W., Wagenmakers, E. J., & Parker, T. H. (2017). Detecting and avoiding likely false-positive findings—a practical guide. *Biological Reviews*, 92, 1941-1968.
- Fortier-Brochu, É., Beaulieu-Bonneau, S., Ivers, H., & Morin, C. M. (2012). Insomnia and daytime cognitive performance: A meta-analysis. *Sleep Medicine Reviews*, 16, 83-94.
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345, 1502-1505.
- Frenda, S. J., Patihis, L., Loftus, E. F., Lewis, H. C., & Fenn, K. M. (2014). Sleep deprivation and false memories. *Psychological Science*, 25, 1674-1681.

- Frey, U., & Morris, R. G. (1997). Synaptic tagging and long-term potentiation. *Nature*, 385, 533-536.
- Fronczek, R., Middelkoop, H. A., Van Dijk, J. G., & Lammers, G. J. (2006). Focusing on vigilance instead of sleepiness in the assessment of narcolepsy: high sensitivity of the Sustained Attention to Response Task (SART). *Sleep*, 29, 187-191.
- Gais, S., Mölle, M., Helms, K., & Born, J. (2002). Learning-dependent increases in sleep spindle density. *Journal of Neuroscience*, 22, 6830-6834.
- Gander, P. H., Marshall, N. S., Harris, R., & Reid, P. (2005). The Epworth Sleepiness Scale: Influence of age, ethnicity, and socioeconomic deprivation. Epworth Sleepiness scores of adults in New Zealand. *Sleep*, 28, 249-254.
- Garbarino, S., Lanteri, P., Durando, P., Magnavita, N., & Sannita, W. (2016). Co-morbidity, mortality, quality of life and the healthcare/welfare/social costs of disordered sleep: a rapid review. *International journal of Environmental Research and Public Health*, 13, 831.
- Goldich, Y., Barkana, Y., Pras, E., Zadok, D., Hartstein, M., & Morad, Y. (2010). The effects of sleep deprivation on oculomotor responses. *Current Eye Research*, 35, 1135-1141.
- Golier, J. A., Yehuda, R., Lupien, S. J., Harvey, P. D., Grossman, R., & Elkin, A. (2002). Memory performance in Holocaust survivors with posttraumatic stress disorder. *American Journal of Psychiatry*, 159, 1682-1688.
- Grandner, M. A., Patel, N. P., Gehrman, P. R., Xie, D., Sha, D., Weaver, T., & Gooneratne, N. (2010). Who gets the best sleep? Ethnic and socioeconomic factors related to sleep complaints. *Sleep Medicine*, 11, 470-478.

- Greathouse, S. M., & Kovera, M. B. (2009). Instruction bias and lineup presentation moderate the effects of administrator knowledge on eyewitness identification. *Law and Human Behavior, 33*, 70-82.
- Groeger, J. A., Zijlstra, F. R. H., & Dijk, D. J. (2004). Sleep quantity, sleep difficulties and their perceived consequences in a representative sample of some 2000 British adults. *Journal of Sleep Research, 13*, 359-371.
- Gronlund, S. D., Carlson, C. A., Dailey, S. B., & Goodsell, C. A. (2009). Robustness of the sequential lineup advantage. *Journal of Experimental Psychology: Applied, 15*(2), 140.
- Gronlund, S. D., Carlson, C. A., Neuschatz, J. S., Goodsell, C. A., Wetmore, S. A., Wooten, A., & Graham, M. (2012). Showups versus lineups: An evaluation using ROC analysis. *Journal of Applied Research in Memory and Cognition, 1*, 221-228.
- Gronlund, S. D., Wixted, J. T., & Mickes, L. (2014). Evaluating eyewitness identification procedures using receiver operating characteristic analysis. *Current Directions in Psychological Science, 23*, 3-10.
- Gudjonsson, G. H. (1984). A new scale of interrogative suggestibility. *Personality & Individual Differences, 5*, 303-314.
- 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.
- Hafner, M., Stepanek, M., Taylor, J., Troxel, W. M., & van Stolk, C. (2017). Why sleep matters—the economic costs of insufficient sleep: A cross-country comparative analysis. *Rand Health Quarterly, 6*.

- Hallgató, E., Győri-Dani, D., Pekár, J., Janacsek, K., & Nemeth, D. (2013). The differential consolidation of perceptual and motor learning in skill acquisition. *Cortex*, 49, 1073-1081.
- Hanert, A., Weber, F. D., Pedersen, A., Born, J., & Bartsch, T. (2017). Sleep in humans stabilizes pattern separation performance. *Journal of Neuroscience*, 37, 12238-12246.
- Hauser, D., Paolacci, G., & Chandler, J. J. (2018). Common concerns with MTurk as a participant pool: Evidence and solutions. Available at <https://psyarxiv.com/uq45c>
- Haw, R. M., & Fisher, R. P. (2004). Effects of administrator-witness contact on eyewitness identification accuracy. *Journal of Applied Psychology*, 89, 1106.
- Hepburn, M., Bollu, P. C., French, B., & Sahota, P. (2018). Sleep medicine: Stroke and sleep. *Missouri Medicine*, 115, 527-532.
- Herscovitch, J., & Broughton, R. (1981). Sensitivity of the Stanford sleepiness scale to the effects of cumulative partial sleep deprivation and recovery oversleeping. *Sleep*, 4(1), 83-92.
- Hirshkowitz, M., Sarwar, A., & Sharafkhaneh, A. (2011). National Sleep Foundation's sleep time duration recommendations: Methodology and results summary. *Sleep Health*, 1, 40-43.
- Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of sleepiness: a new approach. *Psychophysiology*, 10, 431-436.
- Horne, C. M., Watts, A. L., & Norbury, R. (2019). The influence of subjective sleep quality on the association between eveningness and depressive symptoms. *Biological Rhythm Research*, 50, 534-542.



- Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology*, 4, 97-110.
- Hornung, O. P., Regen, F., Danker-Hopfe, H., Heuser, I., & Anghelescu, I. (2008). Sleep-related memory consolidation in depression: an emerging field of research. *Depression and Anxiety*, 25, 163-165.
- Horváth, K., Myers, K., Foster, R., & Plunkett, K. (2015). Napping facilitates word learning in early lexical development. *Journal of Sleep Research*, 24, 503-509.
- Hu, P., Stylos-Allan, M., & Walker, M. P. (2006). Sleep facilitates consolidation of emotional declarative memory. *Psychological Science*, 17, 891-898.
- Hulse, L. M., & Memon, A. (2006). Fatal impact? The effects of emotional arousal and weapon presence on police officers' memories for a simulated crime. *Legal and Criminological Psychology*, 11, 313-325.
- Hutchison, I. C., & Rathore, S. (2015). The role of REM sleep theta activity in emotional memory. *Frontiers in Psychology*, 6, 1439.
- 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, e07903.
- 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, e07903.
- Irwin, M. R., & Vitiello, M. V. (2019). Implications of sleep disturbance and inflammation for Alzheimer's disease dementia. *The Lancet Neurology*, 18, 296-306.

- Izawa, C. (1966). Reinforcement-test sequences in paired-associate learning. *Psychological Reports, 18*, 879-919.
- Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology, 83*, 340-344.
- Jenkins, J. G., & Dallenbach, K. M. (1924). Obliviscence during sleep and waking. *The American Journal of Psychology, 35*, 605-612.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*, 524-532.
- Johns, M. W. (1991). A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep, 14*, 540-545.
- Johns, M., & Hocking, B. (1997). Daytime sleepiness and sleep habits of Australian workers. *Sleep, 20*, 844-847.
- Jones, B. J., Mackay, A., Mantua, J., Schultz, K. S., & Spencer, R. M. (2018). The role of sleep in emotional memory processing in middle age. *Neurobiology of Learning and Memory, 155*, 208-215.
- Joo, S., Baik, I., Yi, H., Jung, K., Kim, J., & Shin, C. (2009). Prevalence of excessive daytime sleepiness and associated factors in the adult population of Korea. *Sleep Medicine, 10*, 182-188.
- Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence–accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1304.

- Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017). An analysis of data quality: Professional panels, student subject pools, and Amazon's Mechanical Turk. *Journal of Advertising*, 46, 141-155.
- Kelemen, E., Bahrendt, M., Born, J., & Inostroza, M. (2014). Hippocampal corticosterone impairs memory consolidation during sleep but improves consolidation in the wake state. *Hippocampus*, 24, 510-515.
- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams Jr, R. B., Alper, S., ... & Batra, R. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, 1, 443-490.
- Klinzing, J. G., Niethard, N., & Born, J. (2019). Mechanisms of systems memory consolidation during sleep. *Nature Neuroscience*, 22(10), 1598-1610.
- Kloepfer, C., Riemann, D., Nofzinger, E. A., Feige, B., Unterrainer, J., O&rsquo, R., ... & Nissen, C. (2009). Memory before and after sleep in patients with moderate obstructive sleep apnea. *Journal of Clinical Sleep Medicine*, 5, 540-548.
- Koen, J. D., & Yonelinas, A. P. (2014). The effects of healthy aging, amnesic mild cognitive impairment, and Alzheimer's disease on recollection and familiarity: A meta-analytic review. *Neuropsychology Review*, 24, 332-354.
- Korman, M., Doyon, J., Doljansky, J., Carrier, J., Dagan, Y., & Karni, A. (2007). Daytime sleep condenses the time course of motor memory consolidation. *Nature Neuroscience*, 10, 1206.

- Korman, M., Raz, N., Flash, T., & Karni, A. (2003). Multiple shifts in the representation of a motor sequence during the acquisition of skilled performance. *Proceedings of the National Academy of Sciences, 100*, 12492-12497.
- Koulack, D. (1997). Recognition memory, circadian rhythms, and sleep. *Perceptual and motor skills, 85*, 99-104.
- Kumar, V. M. (2008). Sleep and sleep disorders. *Indian Journal of Chest Diseases and Allied Sciences, 50*, 129-135.
- 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*, 3-10.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): affective ratings of pictures and instruction manual*. University of Florida, Gainesville. Tech Rep A-8.
- Lau, H. L., Rundek, T., & Ramos, A. R. (2019). Sleep and stroke: New updates on Epidemiology, Pathophysiology, Assessment, and Treatment. *Current Sleep Medicine Reports, 5*, 71-82.
- 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*, 1-13.
- Lerner, I., & Gluck, M. (2019). Sleep and the extraction of hidden regularities: A systematic review and the importance of temporal rules. *Sleep Medicine Reviews, 47*, 39-50.

- Levenson, J. C., Shensa, A., Sidani, J. E., Colditz, J. B., & Primack, B. A. (2017). Social media use before bed and sleep disturbance among young adults in the United States: A nationally representative study. *Sleep, 40*.
- Lewis, P. A., & Durrant, S. J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends in Cognitive Sciences, 15*, 343-351.
- 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*, 2619-2629.
- Lindsay, R. C., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70*, 556-564.
- Liu, Y., Wheaton, A. G., Chapman, D. P., Cunningham, T. J., Lu, H., & Croft, J.B. (2016). Prevalence of healthy sleep duration among adults — United States, 2014. *MMWR Morb Mortal Weekly Report, 65*, 137–141.
- Lo, J. C., Chong, P. L., Ganesan, S., Leong, R. L., & Chee, M. W. (2016). Sleep deprivation increases formation of false memory. *Journal of Sleep Research, 25*, 673-682.
- Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology, 7*, 560-572.
- Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12*, 361-366.
- Loftus, E. F., & Zanni, G. (1975). Eyewitness testimony: The influence of the wording of a question. *Bulletin of the Psychonomic Society, 5*, 86-88.

- Loureiro, F., & Garcia-Marques, T. (2015). Morning or Evening person? Which type are you? Self-assessment of chronotype. *Personality and Individual Differences*, 86, 168-171.
- Luus, C. E., & Wells, G. L. (1991). Eyewitness identification and the selection of distracters for lineups. *Law and Human Behavior*, 15, 43-57.
- Macmillan, N. A., & Creelman, C. D. (2005). Detection theory: A user's guide. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Madigan, S. A., & McCabe, L. (1971). Perfect recall and total forgetting: A problem for models of short-term memory. *Journal of Verbal Learning & Verbal Behavior*, 10, 101-106.
- Majumder, M. S., Cohn, E. L., Mekaru, S. R., Huston, J. E., & Brownstein, J. S. (2015). Substandard vaccination compliance and the 2015 measles outbreak. *JAMA Pediatrics*, 169, 494-495.
- Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of applied Psychology*, 66, 482.
- Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of personality and social psychology*, 13, 330.
- Mander, B. A., Santhanam, S., Saletin, J. M., & Walker, M. P. (2011). Wake deterioration and sleep restoration of human learning. *Current Biology*, 21, 183-184.
- Mantua, J. (2018). Sleep physiology correlations and human memory consolidation: where do we go from here?. *Sleep*, 41.
- Marr, D. (1971). Simple memory: a theory for archicortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 262, 23-81.

- Maslowsky, J., & Ozer, E. J. (2014). Developmental trends in sleep duration in adolescence and young adulthood: evidence from a national United States sample. *Journal of Adolescent Health, 54*, 691-697.
- Massimini, M., Huber, R., Ferrarelli, F., Hill, S., & Tononi, G. (2004). The sleep slow oscillation as a traveling wave. *Journal of Neuroscience, 24*, 6862-6870.
- 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.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and Neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*, 419-457.
- McDougall, W. (1901). Experimentelle Beiträge zur Lehre vom Gedächtniss.
- McHugh, M. L. (2013). The chi-square test of independence. *Biochemia medica: Biochemia Medica, 23*, 143-149.
- Mednick, S. C., Cai, D. J., Shuman, T., Anagnostaras, S., & Wixted, J. T. (2011). An opportunistic theory of cellular and systems consolidation. *Trends in Neurosciences, 34*, 504-514.
- Mickes, L. (2015). Receiver operating characteristic analysis and confidence-accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition, 4*, 93-102.

- Mickes, L. (2016). The effects of verbal descriptions on eyewitness memory: Implications for the real-world. *Journal of Applied Research in Memory and Cognition*, 5, 270-276.
- Mickes, L., & Wixted, J. T. (2015). On the applied implications of the “verbal overshadowing effect”. *Perspectives on Psychological Science*, 10, 400-403.
- Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous versus sequential lineups. *Journal of Experimental Psychology: Applied*, 18, 361-376.
- Mickes, L., Seale-Carlisle, T. M., Wetmore, S. A., Gronlund, S. D., Clark, S. E., Carlson, C. A., ... & Wixted, J. T. (2017). ROC s in Eyewitness Identification: Instructions versus Confidence Ratings. *Applied Cognitive Psychology*, 31, 467-477.
- Miller, A. N., Taylor, S. G., & Bedeian, A. G. (2011). Publish or perish: academic life as management faculty live it. *Career Development International*, 16, 422-445.
- Mitchell, K. J., & Johnson, M. K. (2000). Source monitoring: Attributing mental experiences. *The Oxford handbook of memory*, 179-195.
- Modiguan, V. (1976). Effects on a later recall by delaying initial re- call *Journal of Experimental Psychology: Human Learning & Memory*, 5, 609-622.
- Molinaro, P. F., Arndorfer, A., & Charman, S. D. (2013). Appearance-change instruction effects on eyewitness lineup identification accuracy are not moderated by amount of appearance change. *Law and Human Behavior*, 37, 432.
- Morgan, D. P., Tamminen, J., Seale-Carlisle, T. M., & Mickes, L. (2019). The impact of sleep on eyewitness identifications. *Royal Society Open Science*, 6, 170501.
- Moscovitch, M., & Nadel, L. (1998). Consolidation and the hippocampal complex revisited: In defense of the multiple-trace model. *Current Opinion in Neurobiology*, 8, 297-300.



Müller, G. E. & A. Pilzecker. 1900. Experimentelle Beiträge zur Lehre vom Gedächtnis. *Z. Psychol. Ergänzungsband*, 1 1–300.

Müller, G. E., & Pilzecker, A. (1900). *Experimentelle beiträge zur lehre vom gedächtniss* (Vol. 1). JA Barth.

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Du Sert, N. P., ... & Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, 7, 217-227.

National Research Council. (2014). *Identifying the Culprit: Assessing Eyewitness Identification*,. Washington, DC: The National Academies Press.

National Sleep Foundation. (2013). International Bedroom Poll. Retrieved from <https://www.sleepfoundation.org/professionals/sleep-america-polls/2013-international-bedroom-poll>.

Nemeth, D., Gerbier, E., & Janacsek, K. (2019). Four Pitfalls in Sleep and Memory Research and How to Avoid Them.

Ng, T., Tan, W., 2005. Prevalence and determinants of excessive daytime sleepiness in an Asian multi-ethnic population. *Sleep Medicine* 6, 523–529.

Niethard, N., & Born, J. (2019). Back to baseline: sleep recalibrates synapses. *Nature Neuroscience*, 22, 149-151.

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience*, 14, 1105-1107.

- Nishida, M., Pearsall, J., Buckner, R. L., & Walker, M. P. (2008). REM sleep, prefrontal theta, and the consolidation of human emotional memory. *Cerebral Cortex*, *19*, 1158-1166.
- Nishiyama, T., Mizuno, T., Kojima, M., Suzuki, S., Kitajima, T., Ando, K. B., ... & Nakayama, M. (2014). Criterion validity of the Pittsburgh Sleep Quality Index and Epworth Sleepiness Scale for the diagnosis of sleep disorders. *Sleep Medicine*, *15*, 422-429.
- Norman, Y., Yeagle, E. M., Khuvis, S., Harel, M., Mehta, A. D., & Malach, R. (2019). Hippocampal sharp-wave ripples linked to visual episodic recollection in humans. *Science*, *365*.
- Ohayon, M., Wickwire, E.M., Hirshkowitz, M., Albert, S.M., Avidan, A., Daly, F.J., Dauvilliers, Y., Ferri, R., Fung, C., Gozal, D. & Hazen, N. (2017). National Sleep Foundation's sleep quality recommendations: First report. *Sleep Health*, *3*, 6-19.
- Okado, Y., & Stark, C. E. (2005). Neural activity during encoding predicts false memories created by misinformation. *Learning & Memory*, *12*, 3-11.
- Olaithe, M., Skinner, T. C., Clarke, J., Eastwood, P., & Bucks, R. S. (2013). Can we get more from the Epworth Sleepiness Scale (ESS) than just a single score? *Sleep and Breathing*, *17*, 763-769.
- Pace-Schott, E. F., Germain, A., & Milad, M. R. (2015). Sleep and REM sleep disturbance in the pathophysiology of PTSD: the role of extinction memory. *Biology of mood & Anxiety Disorders*, *5*, 3.

- Pallesen, S., Nordhus, I. H., Omvik, S., Sivertsen, B., Tell, G. S., & Bjorvatn, B. (2007). Prevalence and risk factors of subjective sleepiness in the general adult population. *Sleep, 30*, 619-624.
- Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law and Human Behavior, 36*, 247-255.
- Palmer, M. A., Brewer, N., Weber, N., & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology: Applied, 19*, 55-71.
- Pan, S. C., & Rickard, T. C. (2015). Sleep and motor learning: is there room for consolidation?. *Psychological Bulletin, 141*, 812-834.
- Pardilla-Delgado, E., & Payne, J. D. (2017). The impact of sleep on true and false memory across long delays. *Neurobiology of Learning and Memory, 137*, 123-133.
- Patihis, L., Frenda, S. J., & Loftus, E. F. (2018). False memory tasks do not reliably predict other false memories. *Psychology of Consciousness: Theory, Research, and Practice, 5*, 140-160.
- Payne, J. D., Schacter, D. L., Propper, R. E., Huang, L. W., Wamsley, E. J., Tucker, M. A., ... & Stickgold, R. (2009). The role of sleep in false memory formation. *Neurobiology of Learning and Memory, 92*, 327-334.
- Payne, J. D., Stickgold, R., Swanberg, K., & Kensinger, E. A. (2008). Sleep preferentially enhances memory for emotional components of scenes. *Psychological Science, 19*, 781-788.

- Payne, J. D., Tucker, M. A., Ellenbogen, J. M., Wamsley, E. J., Walker, M. P., Schacter, D. L., & Stickgold, R. (2012). Memory for semantically related and unrelated declarative information: The benefit of sleep, the cost of wake. *PloS One*, 7, e33079.
- Peyrache, A., Khamassi, M., Benchenane, K., Wiener, S. I., & Battaglia, F. P. (2009). Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nature Neuroscience*, 12, 919.
- Pickel, K. L. (1998). Unusualness and threat as possible causes of " weapon focus". *Memory*, 6, 277-295.
- Plihal, W., & Born, J. (1997). Effects of early and late nocturnal sleep on declarative and procedural memory. *Journal of Cognitive Neuroscience*, 9, 534-547.
- Police and Criminal Evidence Act 1984, Code D. 2011. See [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/253831/pace-code-d-2011.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/253831/pace-code-d-2011.pdf) [Accessed 11th March 2017]
- Porter, D., Moss, A., & Reisberg, D. (2014). The Appearance-Change instruction does not improve Line-up identification accuracy. *Applied Cognitive Psychology*, 28, 151-160.
- Press, D. Z., Casement, M. D., Pascual-Leone, A., & Robertson, E. M. (2005). The time course of off-line motor sequence learning. *Cognitive Brain Research*, 25, 375-378.
- Public Health England. (2018). Measles notifications and confirmed cases by oral fluid testing 2013 to 2019 (Q1). Retrieved from <https://www.gov.uk/government/publications/measles-confirmed-cases/measles-notifications-and-confirmed-cases-by-quarter-in-england-2013-to-2015>
- Rahafar, A., Meysam, S. J., Sadeghpour, A., Heidari, Z., & Kasaeian, A. (2015). Psychometric properties of the Persian version of the reduced Morningness-

- Eveningness Questionnaire: Further evidence. *Sleep and Biological Rhythms*, 13, 112-116.
- Rahal, R. M., & Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349, aac4716.
- Rammohan, K. W., Rosenberg, J. H., Lynn, D. J., Blumenfeld, A. M., Pollak, C. P., & Nagaraja, H. N. (2002). Efficacy and safety of modafinil (Provigil®) for the treatment of fatigue in multiple sclerosis: a two centre phase 2 study. *Journal of Neurology, Neurosurgery and Psychiatry*, 72, 179-183.
- Raouf, A. M., Asaad, Y. A., & Al-Hadithi, T. S. (2014). Distribution of chronotypes among a sample of Iraqi Kurdish medical students. *Sultan Qaboos University Medical Journal*, 14, 356-360.
- Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological reviews*, 93, 681-766.
- Rasch, B., Büchel, C., Gais, S., & Born, J. (2007). Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science*, 315, 1426-1429.
- Read, J. D., Yuille, J. C., & Tollestrup, P. (1992). Recollections of a robbery. *Law and Human Behavior*, 16, 425-446.
- Rickard, T. C., Cai, D. J., Rieth, C. A., Jones, J., & Ard, M. C. (2008). Sleep does not enhance motor sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 834-842.
- Riedel, G., Platt, B., & Micheau, J. (2003). Glutamate receptor function in learning and memory. *Behavioural Brain Research*, 140, 1-47.

- Robertson, E. M., Press, D. Z., & Pascual-Leone, A. (2005). Off-line learning and the primary motor cortex. *Journal of Neuroscience*, 25, 6372-6378.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics*, 12, 77.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 803-814.
- Roediger, H.L. & Karpicke, J.D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249-255.
- Rosanova, M., & Ulrich, D. (2005). Pattern-specific associative long-term potentiation induced by a sleep spindle-related spike train. *Journal of Neuroscience*, 25, 9398-9405.
- Rosenthal, L. D., & Dolan, D. C. (2008). The Epworth sleepiness scale in the identification of obstructive sleep apnea. *The Journal of Nervous and Mental Disease*, 196, 429-431.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86, 638-641.
- Rotello, C. M., & Chen, T. (2016). ROC curve analyses of eyewitness identification decisions: An analysis of the recent debate. *Cognitive Research: Principles and Implications*, 1, 10.
- Rothen, N., & Meier, B. (2017). Time-of-day affects prospective memory differently in younger and older adults. *Aging, Neuropsychology, and Cognition*, 24, 600-612.

- Rupp, T. L., Killgore, W. D., & Balkin, T. J. (2010). Socializing by day may affect performance by night: vulnerability to sleep deprivation is differentially mediated by social exposure in extraverts vs introverts. *Sleep*, 33, 1475-1485.
- Ruths, D. (2019). The misinformation machine. *Science*, 363, 348-348.
- 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, 2534-2541.
- Sander, C., Hegerl, U., Wirkner, K., Walter, N., Kocalevent, R. D., Petrowski, K., ... & Hinz, A. (2016). Normative values of the Epworth Sleepiness Scale (ESS), derived from a large German sample. *Sleep & Breathing*, 20, 1337-1345.
- Sauer, J., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence–accuracy relationship for eyewitness identification. *Law and Human Behavior*, 34, 337-347.
- Sauerland, M., Raymaekers, L. H., Otgaar, H., Memon, A., Waltjen, T. T., Nivo, M., ... & Smeets, T. (2016). Stress, stress-induced cortisol responses, and eyewitness identification performance. *Behavioral Sciences & The Law*, 34, 580-594.
- Sauter, C., Popp, R., Danker-Hopfe, H., Büttner, A., Wilhelm, B., Binder, R., ... & Weeß, H. G. (2007). Normative values of the German Epworth sleepiness scale. *Somnologie-Schlafforschung und Schlafmedizin*, 11, 272-278.
- Schmidt, C., Collette, F., Cajochen, C., & Peigneux, P. (2007). A time to think: circadian rhythms in human cognition. *Cognitive Neuropsychology*, 24, 755-789.
- 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.

- Schönauer, M., Alizadeh, S., Jamalabadi, H., Abraham, A., Pawlizki, A., & Gais, S. (2017). Decoding material-specific memory reprocessing during sleep in humans. *Nature Communications*, 8, 1-9.
- 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, 1995-2007.
- Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, 22, 36-71.
- Scott, A. J., Webb, T. L., & Rowse, G. (2017). Does improving sleep lead to better mental health? A protocol for a meta-analytic review of randomised controlled trials. *BMJ open*, 7, e016873.
- Seale-Carlisle, T. M., & Mickes, L. (2016). US line-ups outperform UK line-ups. *Royal Society Open Science*, 3, 160300.
- Seale-Carlisle, T. M., Wetmore, S. A., Flowe, H. D., & Mickes, L. (2019). Designing police lineups to maximize memory performance. *Journal of Experimental Psychology: Applied*, 25, 410-430.
- Semmler, C., Dunn, J., Mickes, L., & Wixted, J. T. (2018). The role of estimator variables in eyewitness identification. *Journal of Experimental Psychology: Applied*, 24, 400-415.
- Shan, Z., Ma, H., Xie, M., Yan, P., Guo, Y., Bao, W., ... & Liu, L. (2015). Sleep duration and risk of type 2 diabetes: A meta-analysis of prospective studies. *Diabetes Care*, 38, 529-537.
- Sheth, B. R., Nguyen, N., & Janvelyan, D. (2009). Does sleep really influence face recognition memory?. *PLoS One*, 4, e5496.



- Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Specification curve: Descriptive and inferential statistics on all reasonable specifications. *Available at SSRN 2694998*.
- Sirota, A., Csicsvari, J., Buhl, D., & Buzsáki, G. (2003). Communication between Neocortex and hippocampus during sleep in rodents. *Proceedings of the National Academy of Sciences, 100*, 2065-2069.
- Skaer, T. L., & Sclar, D. A. (2010). Economic implications of sleep disorders. *Pharmacoeconomics, 28*, 1015-1023.
- Smith, C. N., Wixted, J. T., & Squire, L. R. (2011). The hippocampus supports both recollection and familiarity when memories are strong. *Journal of Neuroscience, 31*, 15693-15702.
- Smith, H. M., & Flowe, H. D. (2015). ROC analysis of the verbal overshadowing effect: Testing the effect of verbalisation on memory sensitivity. *Applied Cognitive Psychology, 29*, 159-168.
- Solomonova, E., Stenstrom, P., Schon, E., Duquette, A., Dubé, S., O'Reilly, C., & Nielsen, T. (2017). Sleep-dependent consolidation of face recognition and its relationship to REM sleep duration, REM density and Stage 2 sleep spindles. *Journal of Sleep Research, 26*, 318-321.
- Spira, A. P., Beaudreau, S. A., Stone, K. L., Kezirian, E. J., Lui, L. Y., Redline, S., ... & Stewart, A. (2011). Reliability and validity of the Pittsburgh Sleep Quality Index and the Epworth Sleepiness Scale in older men. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences, 67*, 433-439.

- Stamatakis, K. A., Kaplan, G. A., & Roberts, R. E. (2007). Short sleep duration across income, education, and race/ethnic groups: population prevalence and growing disparities during 34 years of follow-up. *Annals of Epidemiology*, *17*, 948-955.
- Staresina, B. P., Bergmann, T. O., Bonnefond, M., Van Der Meij, R., Jensen, O., Deuker, L., ... & Fell, J. (2015). Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. *Nature Neuroscience*, *18*, 1679.
- Stark, C. E., Okado, Y., & Loftus, E. F. (2010). Imaging the reconstruction of true and false memories using sensory reactivation and the misinformation paradigms. *Learning & Memory*, *17*, 485-488.
- Stavem, K., Kristiansen, H. A., Kristoffersen, E. S., Kværner, K. J., & Russell, M. B. (2017). Association of excessive daytime sleepiness with migraine and headache frequency in the general population. *The Journal of Headache and Pain*, *18*, 35.
- Stepan, M. E., Dehnke, T. M., & Fenn, K. M. (2017). Sleep and eyewitness memory: Fewer false identifications after sleep when the target is absent from the lineup. *PloS one*, *12*, e0182907.
- Step toe, A., Peacey, V., & Wardle, J. (2006). Sleep duration and health in young adults. *Archives of internal medicine*, *166*, 1689-1692.
- Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, *437*, 1272-1278.
- Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the P value is not enough. *Journal of Graduate Medical Education*, *4*, 279-282.
- Tamaki, M., Matsuoka, T., Nittono, H., & Hori, T. (2008). Fast sleep spindle (13–15 Hz) activity correlates with sleep-dependent improvement in visuomotor performance. *Sleep*, *31*, 204-211.

- Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *Journal of Neuroscience*, 30, 14356-14360.
- Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *Journal of Neuroscience*, 30, 14356-14360.
- Tamminen, J., Ralph, M. A. L., & Lewis, P. A. (2013). The role of sleep spindles and slow-wave activity in integrating new information in semantic memory. *Journal of Neuroscience*, 33, 15376-15381.
- Technical Working Condition for Eyewitness Evidence. 2003 *Eyewitness evidence: A trainer's manual for law enforcement*. Washington, DC: U.S. Department of Justice, Office of Justice Programs. See <https://www.ncjrs.gov/nij/eyewitness/188678.pdf> [Accessed 7th March 2017].
- Teodorescu, M., Consens, F. B., Bria, W. F., Coffey, M. J., McMorris, M. S., Weatherwax, K. J., ... & Chervin, R. D. (2006). Correlates of daytime sleepiness in patients with asthma. *Sleep Medicine*, 7, 607-613.
- Thorley, C. (2013). The effects of recent sleep duration, sleep quality, and current sleepiness on eyewitness memory. *Applied Cognitive Psychology*, 27, 690-695.
- Tononi, G., & Cirelli, C. (2003). Sleep and synaptic homeostasis: a hypothesis. *Brain Research Bulletin*, 62, 143-150.
- Tononi, G., & Cirelli, C. (2014). Sleep and the price of plasticity: from synaptic and cellular homeostasis to memory consolidation and integration. *Neuron*, 81, 12-34.

- Tucker, M. A., & Fishbein, W. (2008). Enhancement of declarative memory performance following a daytime nap is contingent on strength of initial task acquisition. *Sleep, 31*, 197-203.
- 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*, e21737.
- Tulving, E., & Thomson, D. M. (1971). Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology, 87*, 116-124.
- Tunnicliff, J. L., & Clark, S. E. (2000). Selecting foils for identification lineups: Matching suspects or descriptions?. *Law and Human Behavior, 24*, 231-258.
- United States Census Bureau (2019). Educational Attainment in the United States: 2018. Retrieved from <https://www.census.gov/data/tables/2018/demo/education-attainment/cps-detailed-tables.html>
- Urrila, A. S., Stenuit, P., Huhdankoski, O., Kerkhofs, M., & Porkka-Heiskanen, T. (2007). Psychomotor vigilance task performance during total sleep deprivation in young and postmenopausal women. *Behavioural Brain Research, 180*, 42-47.
- Valentine, T., & Mesout, J. (2009). Eyewitness identification under stress in the London Dungeon. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 23*, 151-161.
- Van der Helm, E., & Walker, M. P. (2011). Sleep and emotional memory processing. *Sleep Medicine Clinics, 6*, 31-43.
- Van Ormer, E. B. (1932). Retention after intervals of sleep and waking. *Archives of Psychology, 137*, 49.

- 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.
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7, 632-638.
- 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.
- 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, 679-687.
- Wais, P. E., Squire, L. R., & Wixted, J. T. (2010). In search of recollection and familiarity signals in the hippocampus. *Journal of Cognitive Neuroscience*, 22, 109-123.
- Wais, P. E., Wixted, J. T., Hopkins, R. O., & Squire, L. R. (2006). The hippocampus supports both the recollection and the familiarity components of recognition memory. *Neuron*, 49, 459-466.
- Walker, M. P. (2009). The role of sleep in cognition and emotion. *Annals of the New York Academy of Sciences*, 1156, 168-197.
- Walker, M. P. (2009). The role of slow wave sleep in memory processing. *Journal of clinical sleep medicine: JCSM: Official publication of the American Academy of Sleep Medicine*, 5, 20-26.

- Walker, M. P., & Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron*, 44, 121-133.
- Walker, M. P., & van Der Helm, E. (2009). Overnight therapy? The role of sleep in emotional brain processing. *Psychological Bulletin*, 135, 731-748.
- Walker, M. P., Brakefield, T., Hobson, J. A., & Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature*, 425, 616.
- Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A., & Stickgold, R. (2002). Practice with sleep makes perfect: sleep-dependent motor skill learning. *Neuron*, 35, 205-211.
- Walker, M. P., Stickgold, R., Alsop, D., Gaab, N., & Schlaug, G. (2005). Sleep-dependent motor memory plasticity in the human brain. *Neuroscience*, 133, 911-917.
- Wamsley, E. J. (2019). Memory consolidation during waking rest. *Trends in Cognitive Sciences*, 23, 171-173.
- Watson, N. F., Badr, M. S., Belenky, G., Bliwise, D. L., Buxton, O. M., Buysse, D., Dinges, D. F., ... & Martin, J. L. (2015). Recommended amount of sleep for a healthy adult: A joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society. *Journal of Clinical Sleep Medicine*, 11, 591-592.
- Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). The selection of distractors for eyewitness lineups. *Journal of Applied Psychology*, 78, 835-844.
- Wetmore, S. A., Neuschatz, J. S., Gronlund, S. D., Wooten, A., Goodsell, C. A., & Carlson, C. A. (2015). Effect of retention interval on showup and lineup performance. *Journal of Applied Research in Memory and Cognition*, 4, 8-14.

- Wiebe, S., Carrier, J., Frenette, S., & Gruber, R. (2013). Sleep and sleepiness in children with attention deficit/hyperactivity disorder and controls. *Journal of Sleep Research*, 22, 41-49.
- 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, 1563-1569.
- Wilson, B. M., Seale-Carlisle, T. M., & Mickes, L. (2018). The effects of verbal descriptions on performance in lineups and showups. *Journal of Experimental Psychology: General*, 147, 113.
- Wilson, J. K., Baran, B., Pace-Schott, E. F., Ivry, R. B., & Spencer, R. M. (2012). Sleep modulates word-pair learning but not motor sequence learning in healthy older adults. *Neurobiology of Aging*, 33, 991-1000.
- Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, 265, 676-679.
- Winocur, G., Moscovitch, M., & Bontempi, B. (2010). Memory formation and long-term retention in humans and animals: Convergence towards a transformation account of hippocampal–neocortical interactions. *Neuropsychologia*, 48, 2339-2356.
- Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, 55, 235-269.
- Wixted, J. T., & Carpenter, S. K. (2007). The Wickelgren power law and the Ebbinghaus savings function. *Psychological Science*, 18, 133.
- Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, 117, 1025-1054.

- Wixted, J. T., & Mickes, L. (2012). The field of eyewitness memory should abandon probative value and embrace receiver operating characteristic analysis. *Perspectives on Psychological Science*, 7, 275-278.
- Wixted, J. T., & Squire, L. R. (2004). Recall and recognition are equally impaired in patients with selective hippocampal damage. *Cognitive, Affective, & Behavioral Neuroscience*, 4, 58-66.
- Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, 18, 10-65.
- Wixted, J. T., Vul, E., Mickes, L., & Wilson, B. M. (2018). Models of lineup memory. *Cognitive Psychology*, 105, 81-114.
- Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2001). A field study of own-race bias in South Africa and England. *Psychology, Public Policy, and Law*, 7, 119-133.
- Wu, S., Wang, R., Ma, X., Zhao, Y., Yan, X., & He, J. (2012). Excessive daytime sleepiness assessed by the Epworth Sleepiness Scale and its association with health related quality of life: A population-based study in China. *BMC Public Health*, 12, 84.
- 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*, 88 364-375.
- YouGov. (2018). Most flat earthers consider themselves very religious. Retrieved from <https://today.yougov.com/topics/philosophy/articles-reports/2018/04/02/most-flat-earththers-consider-themselves-religious>



Young, J. L. (1971). Reinforcement-test intervals in paired-associate learning. *Journal of Mathematical Psychology*, 8, 58-81.

