The Role of Sleep in Memory Consolidation: Effects of Lateralisation and Emotion

Chloe Rhianne Newbury

This thesis is submitted for the degree of Doctor of Philosophy

Lancaster University

Department of Psychology

May 2019

Table of Contents

Appendices

Psychology Lancaster University

Postgraduate Research Degrees

FORM: PERMISSION TO

SUBMIT A THESIS IN ALTERNATIVE FORMAT

Name of Candidate: (PLEASE PRINT)	Chloe Rhianne Newbury
Date of Registration:	October, 2015
Registration Number:	31692353
Division and Faculty:	Psychology, Faculty of Science & Technology
Research Degree Title:	The role of sleep in memory consolidation: effects of lateralisation and emotion
Date Thesis due for Submission:	October, 2018

This is to confirm that the above named student has been granted permission to submit a PhD or MPhil thesis in Alternative Format.

Name(s) of Supervisor(s) (PLEASE PRINT):	Padraic Monaghan, Francesca Citron
Signature(s):	P Mon F. Citron

Declaration

I declare that this thesis is my own work completed solely by the author under the supervision of Professor Padraic Monaghan and Dr Francesca Citron, and that it has not been submitted in substantially the same form for the award of a higher degree elsewhere.

A subset of the data analysed in Chapter 4 formed part of my MSc dissertation.

List of Tables

Chapter 2.

Chapter 4.

Chapter 5.

Chapter 6.

List of Figures

Chapter 1.

Chapter 2.

Chapter 3.

Chapter 4.

Chapter 5.

Chapter 6.

List of Abbreviations

AASM	American Association of Sleep Medicine
AAT	Associate Activation Theory
BAS	Backwards Associative Strength
CLS	Complementary Learning Systems
DPT	Dual Process Theory
DRM	Deese-Roediger-McDermott
EEG	Electroencephalography
FAB	Fading Affect Bias
FTT	Fuzzy Trace Theory
iOtA	information Overlap to Abstract
LH	Left Hemisphere
LVF	Left Visual Field
MTT	Multiple Trace Theory
NREM	Non-Rapid Eye Movement
N1	Sleep-Stage 1
N2	Sleep-Stage 2
N3	Sleep-Stage 3
PSG	Polysomnography
PTSD	Post-Traumatic Stress Disorder
REM	Rapid Eye Movement
RH	Right Hemisphere
RT	Response Time
RVF	Right Visual Field
SAM	Self-Assessment Manikin
SFSR	Sleep to Forget, Sleep to Remember
SSS	Stanford Sleepiness Scale
SWS	Slow-Wave Sleep

Acknowledgements

Firstly, I would like to thank my supervisors, Professor Padraic Monaghan and Dr Francesca Citron, for all of their help and guidance throughout the PhD. A special thank you especially to Padraic for supporting me from undergraduate level till now. I have learned so much from you over the past 5 years and have grown so much in confidence. Thank you for all of your advice, I couldn't have asked for a better supervisor. I would also like to thank the members of the lab for being so supportive, I am appreciative of all of your help throughout the PhD. To Rebecca Frost, thank you for all of your help and guidance throughout the years, and for buying me many a Greggs breakfast, they definitely helped! Thank you to John Shaw, my fellow sleep researcher, who kept me sane during late night testing, supplied me with diet Coke, and amused me with Jeremy Kyle stories.

A huge thank you to my partner, Aamer, for being there for me every single day. From the day we met you have been so supportive, and I can't thank you enough for everything you do for me. To my friends, I am so grateful for all of your support. A huge thank you in particular to my best friend, Georgia, for always encouraging me and believing in me, and to Rebecca Iversen, we met on our very first day as nervous and excited PhD students, and I couldn't imagine doing any of this without you.

Finally, I would like to thank my family for everything they do for me. I would not be where I am today without them. My biggest thank you of all goes to my Mum, who has always taught me to believe in myself and to follow my dreams. I appreciate everything that you do for me. Thank you for your continued encouragement, love, and support.

Abstract

Sleep benefits both learning and memory, with offline memory consolidation leading to the reactivation and integration of new information into the long-term store. Previous literature further demonstrates an influence of sleep on memory for related but unseen information, termed false memories. However, current findings provide mixed results regarding sleep's role in the formation of these false memories within the Deese-Roediger-McDermott (DRM) paradigm, with differing memory effects suggested to be due to various task differences. For word recognition, the reactivation of memories during sleep promotes access to the long-term store in the left hemisphere, whereas false memories may lead to equal spread of activation across the two hemispheres. Whether hemispheric processing of memories after sleep affects memories lateralised at encoding, rather than at retrieval, is however unknown. The emotionality of to-be-remembered information also enhances the consolidation effect, with greater veridical memory performance for both negative and positive stimuli. However, the effect of sleep on false memories varying in emotionality is yet unknown. This thesis therefore presents a series of experiments that examine several factors, namely lateralisation and emotionality, that may influence the role of sleep in learning and memory consolidation.

In Chapter 2, we conduct a meta-analysis to establish the influence of sleep on veridical and false memories within the DRM paradigm literature. Although results suggest no overall effect of sleep, whether participants took part in a recall or recognition task, and the number of words within a DRM word list, moderated the effect. The role of sleep in DRM memory consolidation is therefore dependent on specific task features.

Chapter 3 exposes participants to DRM word lists to the left or right hemisphere at encoding. The results demonstrate no hemispheric difference in memory performance after sleep for either veridical or false memories, suggesting that lateralisation effects after sleep are specific to retrieval processes.

In Chapters 4 and 5, we use behavioural testing (Chapter 4) and polysomnography (PSG; Chapter 5) to investigate sleep-dependent veridical and false memory for emotionally negative, positive, and neutral DRM word lists. Findings demonstrate that negative emotion, compared to neutral, enhances veridical memory performance after sleep, whereas wake supports positive memory. Interestingly, sleep spindles during slow wave sleep (SWS) were found to correlate with increased memory performance for emotional words. In Chapter 6, we investigate the effect of sleep on learning of and memory for novel metaphorical word pairs of either negative, positive, or neutral valence. Findings demonstrate increased memory for emotional word pairs after sleep than wake.

These results help clarify the role sleep plays in the formation of emotional memories, and highlights factors that modulate the effect of sleep on both veridical and false memories. We provide evidence that lateralisation effects are specific to retrieval processes, and suggest that sleep boosts the consolidation of emotional information, indicating a potential role of sleep spindles specific to SWS in the consolidation of emotional memories.

1. Literature Review

1.1 General Introduction to the Thesis

Sleep is associated with improvements in motor skill learning (Walker, Brakefield, Morgan, Hobson, & Stickgold, 2002), problem solving (Sio, Monaghan, & Ormerod, 2013) the acquisition of language (Frost & Monaghan, 2017), and the integration of new information (Tamminen, Lambon, Ralph, & Lewis, 2013). In particular, sleep has been found to enhance declarative memories, with literature suggesting a role of hippocampal and neocortical systems in the reactivation and integration of memories from a short-term store into long-term memory (Marshall & Born, 2007; Rasch & Born, 2013). However, limited capacity within the long-term store means that specific details of memories are not consolidated, but rather memories are integrated such that the general theme or “gist” of the memory is stored for later retrieval. Thus, although sleep has been found to enhance veridical (accurate) memory performance, research also indicates a role of sleep for unseen, related information, termed false memories (Diekelmann, Born, & Wagner, 2010; Payne et al., 2009). However, inconsistencies emerge, with some studies showing no effect or a reduction in false memories after sleep (Diekelmann, Landolt, Lahl, Born, & Wagner, 2008; Fenn, Gallo, Margoliash, Roediger, & Nusbaum, 2009). These inconsistencies in memory performance may be due to a number of factors, some of which will be explored throughout this thesis. For example, research suggests that there are hemispheric asymmetries in the retrieval of previously consolidated memories, with evidence for a left hemisphere (LH) bias for veridical memory for words (Monaghan, Shaw, Ashworth-Lord, & Newbury, 2017), and a right hemisphere (RH) bias for false memories (Shaw & Monaghan, 2017). Furthermore, sleep has been found to prioritize the consolidation of emotional memories (Rasch & Born, 2013; van der Helm & Walker, 2009), however specific effects of valence and arousal are as of yet unknown.

The aim of this thesis is thus to examine the influence of sleep on both veridical and false memories, and further explore the influences of lateralisation and emotionality on memory consolidation. In this chapter, I will first provide an overview of current literature on sleep and memory separately, before discussing key theories of sleep's influence on memory. I will then address the current literature on hemispheric asymmetries and emotionality in memory and their interactions with sleep, and highlight the key questions that this thesis aims to address.

1.2 Sleep

Sleep is characterised by a resting state of inactivity and a lack of consciousness. Sleep is essential for survival and is a fundamental part of the body's circadian rhythm, however efforts to understand the functions of sleep for humans is still in progress. From the use of polysomnography (PSG), it is clear that despite human's physical inactivity during sleep, the brain remains in an active state. From such methods, we know that sleep is split by two main stages; rapid-eye-movement (REM), and non-REM (NREM). NREM sleep can be further categorised into stage 1 (N1), stage 2 (N2), and stage 3 (N3), classified by the American Association of Sleep Medicine (AASM).

Sleep is divided into sleep cycles lasting approximately 90-minutes that usually contain each stage of sleep, cycling from NREM to REM throughout each cycle (Walker, 2009; Figure 1). However, each cycle does not include an equal amount of each sleep stage. N3, otherwise known as slow-wave-sleep (SWS) has been found to occur predominantly in the first half of a night of sleep, and is characterised by slow, high amplitude EEG oscillations and delta waves (Mölle, Marshall, Gais, & Born, 2002). In contrast, REM sleep occurs to a greater degree in the latter half of a night of sleep, and generally consists of fast, low amplitude oscillations (Born, Rasch & Gais, 2006). NREM stage 2 sleep takes up approximately 50% of total sleep time, and consists of waxing and waning sleep spindles (characterised by 11-15 Hz oscillations lasting up to 3s) and k-complexes (a negative peak transitioning into a positive wave). Finally, N1 takes up only a small portion of the amount of sleep time (3-8%) and is characterised by theta waves and slow rolling of the eyes.

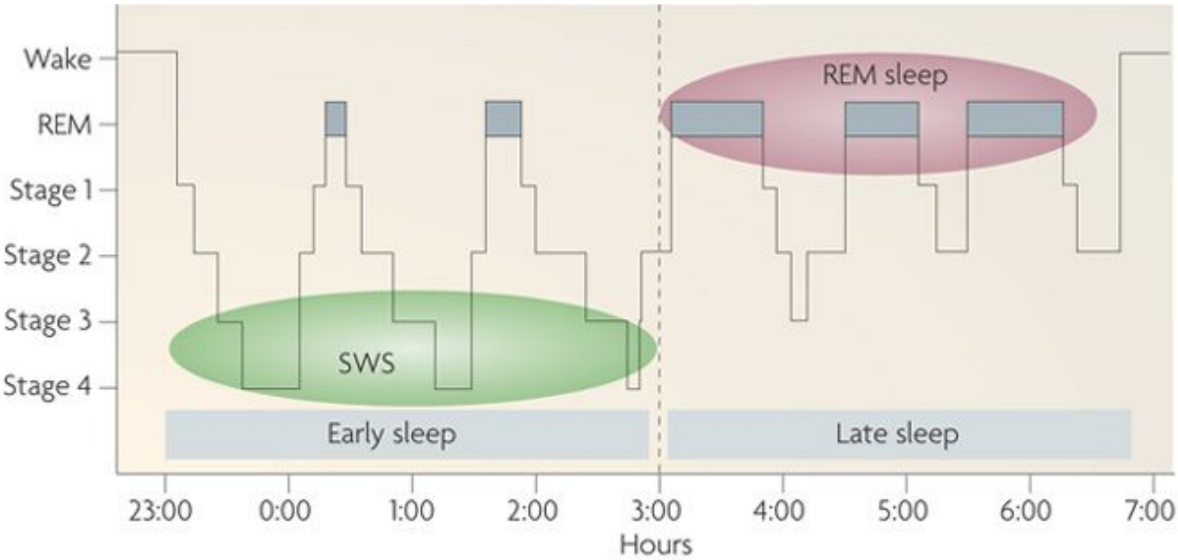


Figure 1.1. The sleep cycle throughout a night of sleep. Earlier periods of sleep are rich in SWS, whereas later periods contain greater amounts of REM sleep. Source: (Diekelmann & Born, 2010).

1.3 Memory Consolidation

There are two fundamental memory systems: declarative memory and procedural memory (Squire & Zola, 1996). Declarative memory is considered memory for facts that we are able to consciously access (i.e., knowing “what”). Declarative memory can be further sub-categorised to include memory for events or information from the past (episodic memory), and memory for general knowledge (sematic memory) (Tulving, 1985). Declarative memories are thought to be hippocampus dependent (Eichenbaum, 2000), with evidence indicating a role of the hippocampus for a range of memory processes, including encoding, short and long-term memory consolidation, and retrieval (for a review see Riedel & Micheau, 2001). On the other hand, procedural memory consists of memory for knowing "how", such as learning of skills and implicit knowledge, and is less reliant on hippocampal structures.

There is not one specific time period or event that creates a memory; a memory is the culmination of encoding, consolidation, and retrieval processes that occur over time. During encoding of a memory, information is stored in the short-term memory store within hippocampal memory systems, and these memories are highly susceptible to forgetting and interference. After information has been encoded, memory consolidation can occur, which is typically the process in which memory becomes less susceptible to interference over time (McGaugh, 2000). This period of consolidation can take place both during wakefulness or sleep. Finally, retrieval processes allow for the recall or recognition of consolidated memories.

1.3.1 Theories of Memory Consolidation

Memory consolidation as a concept was first proposed by Müller and Pilzecker (1900) (Lechner, Squire, & Byrne, 1999), who suggested that during encoding memories are susceptible to interference, and it is only after a period of consolidation can memories become strengthened. Evidence of memory consolidation had already been proposed by Ribot (1882), who observed patients with retrograde amnesia following a brain injury, who could recall events from the past (e.g., childhood events), but memories that were encoded closer to their injury were easily forgotten. More recently, a number of theories of memory consolidation have been discussed throughout the literature. First, the Complementary Learning Systems (CLS) hypothesis was proposed. This hypothesis was developed from ideas taken from Marr (1971) who was one of the first to suggest a model of memory being "transferred" from hippocampal systems into neocortical systems, and from Squire, Cohen and Nadel (1984), who suggested that consolidation required interaction between hippocampal and extrahippocampal (neocortical) circuits. Later, to account for limitations of an early CLS hypothesis, Nadel and Moscovitch (1997) proposed a Multiple Trace Theory (MTT) account of memory consolidation, to consider differences between episodic and semantic memories.

1.3.1.1 Complementary Learning Systems Hypothesis

The Complementary Learning Systems hypothesis (McClelland, McNaughton & O'Reilly, 1995), derived from the standard model of memory consolidation, suggests that the hippocampus plays a role in the encoding and storage of newly learned information, which is then transferred to extrahippocampal, neocortical sites during periods of consolidation. The theory proposes "fast learning" of information that is encoded and stored in the hippocampus, and that the hippocampus gradually transfers information to the neocortex, termed "slow learning". This period of slow learning allows for the assimilation of new information into the long-term store without catastrophic interference (such that prior knowledge is not forgotten at the expense of new information). Thus, over time connections between neocortical areas become stronger, and the connection between hippocampal and neocortical areas becomes weaker, until the hippocampus is no longer necessary for storage of information (see Figure 2). More recently, Kumaran, Hassabis, and McClelland

(2016) updated this model, proposing that replay of memories within the hippocampus can also support memory through generalisation and reasoning, and that "fast learning" can occur in the neocortex if to-be-remembered information is consistent with prior knowledge. Thus, rather than their earlier model of distinct learning systems specific for fast and slow learning, the CLS is now updated to support a more interleaved system between hippocampal and neocortical networks.

Evidence for CLS primarily comes from patients with brain lesions. For example, Scoville and Milner (1957) found that in patient HM, removal of brain lesions in hippocampal areas was related to retrograde amnesia, such that loss of hippocampal networks was related to loss of declarative memory. However, Scoville and Milner point out that lesions were removed from both the hippocampus and the amygdala, and thus effects on short term memory cannot merely be attributed to the removal of hippocampal regions. More recently however, support for CLS comes from findings of greater activity in the hippocampus than neocortical systems during immediate recall of memories, whereas over time neocortical systems played a greater role in memory retrieval, and the connection between hippocampal and neocortical systems was reduced (Takashima et al., 2006). Similarly, evidence of greater consolidation of hippocampus dependent arbitrary form meaning mappings over systematic mappings (that can be processed neocortically) supports a CLS hypothesis of a strengthening of hippocampal and neocortical connections over time (Mirković & Gaskell, 2016). Furthermore, computational models of learning and memory provide evidence for a fast learning hippocampal system, and slow learning in the neocortical systems (for a review see O'Reilly & Norman, 2002).

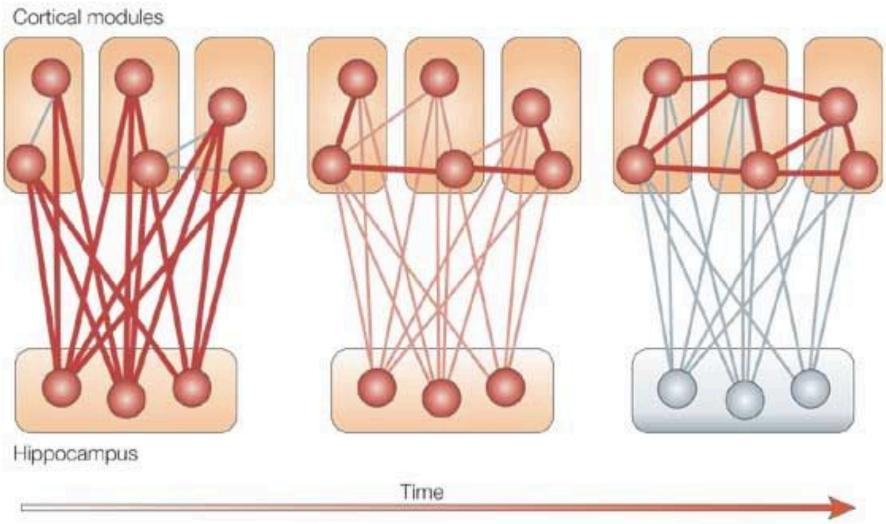


Figure 1.2. A model of Complementary Learning Systems. The model proposes that memories are initially stored in hippocampal systems. As memories are reactivated and integrated over time, they are strengthened within neocortical systems, and hippocampal systems become weaker until such memory representations are no longer needed. Source: (Frankland & Bontempi, 2005).

1.3.1.2 Multiple Trace Theory

Despite evidence for a CLS account of memory consolidation, weaknesses of the theory were evident. For example, the CLS hypothesis does not differentiate between episodic and semantic memory consolidation. For episodic memories, events become less vivid and more gist-based over time (Levine, Svoboda, Hay, Winocur, & Moscovitch, 2002), until they eventually become more semantic or fact-based. Thus, Nadel and Moscovitch (1997) proposed a Multiple Trace Theory to account for these differences in declarative memory. MTT proposes that consolidation of memory is dependent on whether such explicit memory is episodic or semantic. Semantic memories are treated in a similar way to that suggested in CLS, whereby consolidation from hippocampal to extrahippocampal systems allows for the strengthening of memory and the integration of information into the long-term store.

In contrast, episodic memories are thought to be dependent upon hippocampal systems in order to retrieve detailed information of the event. Evidence of dissociations between episodic and semantic memory in patients with retrograde amnesia provides support for MTT. For example, Moscovitch et al. (2005) provide a review of the effects of the hippocampus for memory retention and retrieval in those patients with lesions, and show that hippocampal areas and medial temporal lobes are both required for the retrieval of episodic memories specifically. Furthermore, neuroimaging results indicate that the hippocampus is activated when retrieving memories that are as old as 20-30 years, suggesting a necessity for the hippocampus even for long-term memory (Nadel, Samsonovich, Ryan, & Moscovitch, 2000).

In conclusion, both the updated CLS hypothesis and the MTT propose that hippocampal and neocortical systems are actively involved in the consolidation of memories in a complementary nature. In an effort to merge the two theories together, Winocur and Moscovitch (2011) therefore proposed a transformation model of memory consolidation. As such, declarative memories are reorganised over time, suggesting that as the neural networks involved in a memory representation change, so does the nature of the memory. Much like MTT, they suggest that episodic memories that are dependent on hippocampal systems become semantic and gist-based through repeated reactivation, at which point they are integrated into neocortical structures. Hippocampal networks are not needed for the retrieval of these gist-based memory traces, however retrieval of episodic memories will require the hippocampus for as long as such memories are retained.

1.3.2 False Memories

The consolidation of memories from hippocampal based, context-dependent information to more semantic, gist-based memories in the neocortex supports long-term memory. However, it has also been suggested that such gist-based memory representations may increase the likelihood of misremembering, termed “false memories”. Such distortions and errors in memory occur often (Addis, Wong, & Schacter, 2007; Schacter & Slotnick, 2004). One way to test susceptibility to false memories within a lab based setting is the Deese-Roediger-McDermott (DRM) word lists paradigm (Roediger & McDermott, 1995). Within the DRM paradigm, participants are exposed to a number of word lists, with each list consisting of thematically related words (e.g., *dream, bed, night, snooze etc*.). During subsequent testing, participants are asked to either recall or recognise words from the initial lists. Words are categorised as either those previously seen in the initial word lists (old words), words that did not appear in the initial lists and are unrelated to the words in the lists (new words), and words that did not appear in the initial lists but are related to the word lists (unseen, related lure words, e.g., *sleep* for the above list). Findings indicate high false recall and recognition of previously unseen lure words with related meanings (McDermott, 1996; Roediger & McDermott, 2000).

Two main theories have been proposed that provide an explanation for the false recollection of these unseen related lure words from memory. Associative activation theory (AAT) suggests that unseen, related lure words are activated due to their semantic association with the initially encoded word lists (Arndt & Reder, 2003; Reder, Park, & Kieffaber, 2009). If this theory is accurate, then we would expect to see an increase in unseen related lure word acceptance for those words that are more strongly associated to the initial word lists, termed higher in backwards associative strength (BAS). This is supported by a number of studies (Gallo & Roediger, 2003; Howe, Wimmer, Gagnon, & Plumpton, 2009; Roediger, Watson, McDermott, & Gallo, 2001), suggesting that unseen, related lure words are more greatly activated if their association to the initial word lists is greater.

However, a second theory, fuzzy trace theory (FTT), suggests that gist memory traces play a role in the development of false memories (Brainerd, 2002; Brainerd, Reyna, & Ceci, 2008; Reyna & Brainerd, 1995). This theory suggests that there are both verbatim memory traces, which are memories for the specific detail of the memory, as well as gist memory traces, that support memory for the general meaning of the information or the overall theme. Applying FTT to the DRM paradigm, it is suggested that these verbatim memory traces aim to successfully reject related lure words, whereas gist memory traces may lead to false acceptance of the lure words due to remembering the general theme of a word list (Reyna, Corbin, Weldon, & Brainerd, 2016).

1.3.2.1 Moderators that Influence False Memories

Despite a large body of literature supporting this false memory effect within the DRM paradigm, there are a number of factors that have been found to influence the effect. Thus, the way that we encode, consolidate, and retrieve memories is largely dependent on task-specific features. Some of these features are outlined in the following section.

Recall versus Recognition Testing

The activation-monitoring framework (Collins & Loftus, 1975; Johnson, Hashtroudi, & Lindsay, 1993) proposes that during retrieval of memories, monitoring processes work to locate the source of a memory and activate that memory if necessary. During tests of recognition, words presented to participants act as monitoring cues, aiding in source monitoring by allowing for the active rejection of previously unseen words. In contrast, during tests of free recall no monitoring cues are available, and so a larger number of associated words are activated, leading to increases in false memories (Watson, McDermott, & Balota, 2004). A number of studies using the DRM paradigm support this, finding that participants are more susceptible to false memories in tests of recall than in tests of recognition (El Sharkawy, Groth, Vetter, Beraldi, & Fast, 2008; Howe, Candel, Otgaar, Malone, & Wimmer, 2010).

Interval Between Encoding and Retrieval

Various types of interval times have been used in studies investigating false memories, with some studies testing false recollection immediately after the training phase, whilst others adopt a short delay filled with a distractor task (such as maths questions). McDermott (1996; Experiment 1) found no difference in performance on a false recollection task between participants who were tested immediately after learning and those given a short delay. However, it is suggested that longer delays between initial exposures to to-be-remembered lists and testing would lead to differences in false memory rates. As memories become integrated into neocortical systems over time, memory representations become more gist-based (FTT), and associations with prior knowledge are strengthened (AAT). In support of this, McDermott (1996; Experiment 2) reported that a 2-day delay between training and testing led to higher recall of unseen, related lure words than previously studied items, demonstrating the robustness of false memories over time in comparison to veridical memory. Similarly, Payne, Elie, Blackwell, and Neuschatz (1996) found that false recognition was much more persistent than veridical memory after a delay of just one day. These findings support theories of memory consolidation, such that gist-based neocortical representations are much more easily accessed over time than specific, context-dependent hippocampal representations. Likewise, findings suggest that false recall and recognition does not decline at such a significant rate as memory for veridical memory after one week (Thapar & McDermott, 2001) or 3 weeks (Toglia, Neuschatz, & Goodwin, 1999), again supporting the idea that false memories are much more persistent over time.

Despite this, there are some variations in the literature as to the persistence of false memories. Contrary research has found that false recognition reduced at a similar rate to veridical recognition over delays ranging from two days to one week (Brainerd, Wright, Reyna, & Mojardin, 2001; Lampinen & Schwartz, 2000; Neuschatz, Payne, Lampinen, & Toglia, 2001). Nevertheless, the interval between encoding and subsequent retrieval should be considered as a potential moderator that may lead to variations in false memory performance.

Presentation of Stimuli

Studies investigating false memories using the DRM paradigm differ in terms of the method in which stimuli are presented at encoding. Previous research has indicated that there are differences in false memory rates following visual compared to auditory presentation of stimuli. Smith and Hunt (1998) compared false recall and recognition of word lists presented visually and auditorily and found that visual presentation significantly reduced the false memory effect in tests of recall and recognition. In tests of recall, Kellogg (2001) studied the effects of presenting stimuli visually versus auditorily, and the effects of written versus spoken recall, on false recollections. It was found that false recall was higher for those stimuli presented auditorily compared to visually, but only for written recall as opposed to spoken recall. Similarly, Smith and Engle (2011) found that, for participants of higher working memory capacity, false recall was reduced when viewing visual presentations of the stimuli, compared to auditory presentation. There was no such difference in modality for participants in the low working memory span condition. Thus, visual presentation of stimuli may lead to a reduction in susceptibility to false memories, although this effect may not occur for all participants. This difference in presentation modalities should therefore be considered when comparing studies of false memories that use the DRM paradigm.

Presentation Duration

A number of studies have highlighted a potential influence of presentation duration on the development of false memories, however the results lack consistency. McDermott and Watson (2001) found that for short presentation durations of up to 250ms, false recall increased with increasing presentation duration. However, when participants viewed words for a longer time period of 1000-5000ms, false recall actually declined as presentation time increased. This U-shaped pattern was further highlighted by Smith and Kimball (2012), however the effect of presentation duration was not as strong for delayed false recall, with evidence of high levels of false recall for all variations of fast presentation duration.

For false recognition, evidence suggests an increase in false memory rates when presentation duration is increased. Seamon, Luo, and Gallo (1998) compared false recognition when words were presented for 20ms versus 250ms and found no difference in false recognition rates. If participants were exposed to stimuli for 2000ms they showed a significant increase in false recognition compared to presentation of 20ms. However, it is not known whether false recognition would be susceptible to a similar inverted-U type pattern seen for false recall, with false memories declining for longer presentation durations of up to 5000ms. Further research is needed to investigate whether presentation duration has a significant impact on false recognition in the same way as false recall, and should be considered when comparing research that investigates such false memory effects.

1.3.2 Conclusions

Both the CLS hypothesis and MTT suggest an interaction between hippocampal representations, whereby memories are encoded with their specific detail, and neocortical representations, that store more general, semantic information that is more gist based with broader associations. This interaction between hippocampal and neocortical systems thus allows for stronger veridical memories, but also increases the formation of false memories, either through increasing associations between new and prior knowledge, or through gist-based memory representations. However, a large body of literature suggests that although memories can be consolidated during wakefulness, sleep plays an optimal role in the integration of memories into neocortical systems. The following section will discuss previous findings of effects of sleep on both veridical and false memories, and outline theories proposed to account for this specific role of sleep on memory.

1.4 Sleep's Role in Memory Consolidation

Memory for newly learned information benefits from the process of consolidation specifically during off-line periods of sleep, with evidence of a beneficial effect of sleep for the consolidation of both declarative and procedural memories (Marshall & Born, 2007; Rasch & Born, 2013). During sleep, theories of memory consolidation predict that memories initially encoded in the hippocampus are integrated into extrahippocampal regions, and hippocampal population bursts in particular allow for the integration of newly acquired memories to the neocortex (Buzsaki, 1998). During sleep, there is evidence of increased activity of hippocampal cells that were also active during encoding (Pavlides & Winson, 1989; Wilson & McNaughton, 1994), and findings indicate a significant correlation between activity in the hippocampus and neocortical areas during sleep (Qin, McNaughton, Skaggs, & Barnes, 1997; Sirota, Csicsvari, Buhl, & Buzsáki, 2003).

Behavioural studies of sleep and memory support a benefit of sleep for declarative memory consolidation. For example, memory for word pairs has been found to be greater after a delay including a period of sleep compared to the same time delay during daytime wakefulness (Plihal & Born, 1997; Wilson, Baran, Pace-Schott, Ivry, & Spencer, 2012). Learning of novel information has also been found to be enhanced after a period of sleep; Dumay and Gaskell (2007) found that recognition time of newly learned novel words significantly increased after sleep, suggesting that it took longer for participants to differentiate between whether the stimulus was a newly learned word or a word already in their lexicon. Thus, new information was integrated into prior knowledge during a period of sleep, whereas those who had a period of wakefulness did not show this same consolidation of novel words. Similarly, Davis, Di Betta, Macdonald and Gaskell (2009) found increased performance accuracy during a word recollection task of newly learned words after sleep than wake. Using functional magnetic resonance imaging (fMRI), they found evidence for increased activation in the hippocampus during word learning that was significantly linked to increased retention of the novel words, and there was evidence of similar neocortical activation in response to both newly learned and already acquired words, suggesting that those words had been consolidated into the long-term store. This study further supports the hypothesis that our memories are first stored in the hippocampus, and during sleep are integrated into the long-term store in the neo-cortex.

Within the literature there are various theories to account for sleep-dependent memory consolidation, namely the dual process theory, the sequential hypothesis, and the active systems consolidation hypothesis.

1.4.1 Theories of Sleep and Memory

1.4.1.1 Dual Process Theory

The Dual Process Theory (DPT) suggests that SWS is particularly beneficial for the consolidation of declarative memories, whereas procedural memories benefit from REM sleep (Gais & Born, 2004; Marshall & Born, 2007). Studies investigating DPT have used the night-half paradigm, whereby encoding of memories occurs before either the first night-half, or before the second night-half. Since SWS occurs to a greater extent in the first half of a night of sleep, whereas REM sleep is more dominant in the second half of sleep, the paradigm allows for an understanding of the effects of the two stages of sleep on memory. Findings from night-half studies support the DPT, with results revealing a specific benefit of SWS for declarative memory tasks (Fowler, Sullivan, & Ekstrand, 1973; Plihal & Born, 1997; Yaroush, Sullivan, & Ekstrand, 1971). Stimulating slow oscillations during SWS has also been found to significantly increase memory performance in a word-pair learning task (Marshall, Helgadóttir, Mölle, & Born, 2006), and SWS is likely to facilitate declarative memory even after a short nap (Tucker et al., 2006). In contrast, findings support a role of REM sleep for procedural memory tasks such as mirror tracing (Plihal & Born, 1997) and priming (Wagner, Hallschmid, Verleger, & Born, 2002).

Despite evidence for the DPT in terms of the differential effects of REM and SWS on memory, findings also suggest a potential effect of NREM sleep on procedural memory (Ackermann & Rasch, 2014; Gais, Plihal, Wagner, & Born, 2000). A major criticism of the DPT is that it does not take into account potential influences of other components of sleep occurring during NREM, such as stage 2 sleep and subsequent sleep spindles. Stage 2 sleep and sleep spindles within stage 2 have been found to correlate with both procedural and declarative memory consolidation (Clemens, Fabó & Halász, 2006; Fogel & Smith, 2006; Gais, Mölle, Helms & Born, 2002; Peters, Ray, Smith, & Smith, 2008; Schmidt et al., 2006), and thus should be considered within any account of sleep-dependent memory consolidation.

1.4.1.2 The Sequential Hypothesis

The sequential hypothesis postulates that both NREM (SWS) and REM sleep are involved in memory consolidation, and that the cyclic nature of these sleep stages in particular is important for the processing of memories (Ambrosini & Giuditta, 2001; Giuditta, Ambrosini, Montagnese, Mandile, Cotugno, Grassi Zucconi, & Vescia, 1995). During SWS, selective memories are strengthened, and then during REM sleep these memories are integrated into neocortical stores. Mednick, Nakayama, and Stickgold (2003) provide support for the sequential hypothesis, with findings indicating that a 90-minute nap containing both NREM and REM sleep led to task improvements, whereas a 60-minute nap containing only NREM sleep did not. Similarly, Gais et al. (2000) found that a night of sleep containing both SWS-rich sleep in the first half of the night and REM-rich sleep in the second night-half led to increases in memory performance compared to a night of sleep containing only SWS-rich or REM-rich sleep. Thus, it is suggested that both SWS and REM sleep are required for the consolidation of memories.

However, similarly to DPT, the sequential hypothesis neglects the effects of N2 and sleep spindles in the integration of memories into neocortical systems. Despite this, current research is beginning to recognise the potential influence of sleep spindles on memory consolidation. Antony, Schönauer, Staresina, and Cairney (2018) propose a theory of memory reinstatement that is dependent on the sequential process of sleep spindles. Specifically, they suggest that the reactivation of memories occurs during spindle events, whereas subsequent periods of sleep spindle refractoriness (a period of sleep immediately after a sleep spindle when other spindles are unlikely to occur) support the processing of such newly acquired memories. They further postulate that sleep spindle refractoriness supports the strengthening of memories and reduces interference from unrelated information. This is supported by recent evidence suggesting an important role in the timing of sleep spindles during memory consolidation (Antony, Piloto, Wang, Pacheco, Norman, & Paller, 2018), however further research is needed to more greatly explore the role of sleep spindles in memory reactivation and consolidation.

1.4.1.3 Active Systems Consolidation

The active systems consolidation hypothesis focuses on the neural mechanisms involved in the integration of declarative memories into neocortical stores, as opposed to specific REM versus NREM influences. As with models of memory consolidation, the active systems consolidation hypothesis proposes that memories are initially encoded and stored in hippocampal and neocortical networks. Reactivation of memories during SWS supports the gradual strengthening of neocortical networks, leading to the reorganisation of the memory representation. In particular, slow oscillations during SWS, along with hippocampal sharp wave-ripples and sleep spindles, allow for the repeated reactivation of such memories within the hippocampus. The stabilizing of these memories is then hypothesised to take place during successive periods of REM sleep.

Evidence in support of active systems consolidation comes from studies of memory reactivation during SWS. For example, Rasch, Buchel, Gais, and Born (2007) found that odour cueing during post-learning SWS led to the reactivation and consolidation of declarative memories, and Antony, Gobel, O'Hare, Reber, and Paller (2012) found similar results for skill learning. Furthermore, Marshall et al. (2006), using transcranial stimulation, found that applying slow oscillating stimulation induced SWS and sleep spindle activity in the frontal cortex, which correlated with enhanced memory performance. In contrast, stimulations at a frequency band similar to that of REM sleep had no effect on declarative memory performance.

1.4.2 Sleep's Role in False Memories

The reactivation and integration of new memories into the long-term store not only leads to increases in veridical memory performance, but may also enhance false memories for unseen, related information. If sleep supports the integration of memories from context dependent hippocampal systems to neocortical, context independent representations either through a broader spread of activation of a number of associations, or greater memory for the gist, then we would expect to see increases in DRM false memories. See Chapter 2 for a meta-analytical review of previous studies employing the DRM paradigm to examine the role of sleep in both veridical and false memory consolidation. Within the meta-analysis, we outline potential moderating factors on sleep-dependent veridical and false memories, suggesting that sleep’s role in the consolidation of veridical and false memories is dependent on various task constraints, namely recall versus recognition testing and effects of list length.

The influence of specific sleep architecture on the formation of false memories remains unanswered. Such as active systems consolidation indicates a role of SWS in the reactivation and integration of veridical memories, SWS has also been linked to the formation of memory for the gist, due to abstraction and integration of memories with distantly related information. Lewis and Durrant (2011) proposed the information overlap to abstract (iOtA) model, arguing that reactivation of memories during SWS not only strengthens veridical memory, but also leads to strengthening of shared schemas, leading to gist-based memory representations. This theory supports FTT as an explanation for the formation of false memories, with initial memories being encoded in the hippocampus, and during reactivation these memories are integrated into gist-based memory traces.

Despite the suggestion that SWS may play a role in gist-based memory representations, there is little evidence to support such an influence of SWS on the formation of false memories. Diekelmann, Büchel, Born, and Wagner (2011) found that increases in SWS correlated with a reduced susceptibility to interference, thus SWS may actually protect against the formation of DRM false memories. Indeed, Pardilla-Delgado and Payne (2017) found that SWS negatively correlated with false memories (only in low performing participants).

Sleep spindles have also recently been implicated in the formation of false memories. Shaw and Monaghan (2017) reported a correlation between increases in

RH sleep spindles and DRM false memories. Sleep spindles have been previously suggested to support the integration of new information into semantic memory (Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010; Tamminen, et al., 2013), thus sleep spindles may play a distinct role in the consolidation of memories, either by increasing the associations between memories, or supporting the abstraction of the gist. Again, further evidence is needed to fully understand the role of sleep spindles in the integration of memories from hippocampal to neocortical systems.

1.4.3 Conclusions

The active systems consolidation hypothesis provides the current account of how sleep-dependent memory consolidation occurs, with reactivation of memories occurring during SWS, and the stabilization of such memories occurring during REM sleep. However, despite strong evidence for a role of sleep in veridical memories, there are also suggestions for an influence of sleep on false memories, though specific sleep architecture involved in the formation of these false memories is still unclear. Whether sleep supports memory for unseen, related information, through either associative activation or gist-based memory traces, is dependent on various features of the memory task outlined in Chapter 2, and thus this may explain a lack of a clear influence of sleep and sleep architecture on the formation of false memories.

1.5.1 Hemispheric Processing of Memories

Research has indicated a LH advantage in tests of language processing and word learning (Ellis, 2004; Pirozzolo & Rayner, 1977), whereby stimuli are presented to either the right visual field (RVF) resulting in LH involvement, or the left visual field (LVF) resulting in RH involvement. However, both the LH and RH have been found to be involved in the processing of language in distinctly different ways. The

LH processes words that are more closely related, due to narrow activation of distinct networks. On the other hand, the RH activates more broad semantic networks with weaker related meanings of associated words (Beeman & Bowden, 2000; Beeman & Chiarello, 1998; Burgess & Simpson, 1988; Monaghan, Shillcock, & McDonald, 2004).

1.5.2 Hemispheric Processing of False Memories

According to theories of false memories (AAT, FTT), the reactivation of memories leads to broader semantic networks, generating either increased semantic associations or greater gist-based memory representations. Evidence suggests a bias for false memories specifically in the RH, consistent with activation of broader semantic networks within the RH. Studies of lateralisation effects on false memories generally present DRM word lists centrally, and during retrieval target words are presented to either the LVF or RVF. For veridical memories, Ito (2001) found that participants were more likely to accurately discriminate old words from lure words when they were presented to the LH in comparison to the RH, supporting a LH advantage for activating a narrow range of semantic associates. For false memories, Westerberg and Marsolek (2003) and Bellamy and Shillcock (2007) found that lure words were correctly rejected in the LH significantly more than in the RH. Participants had greater difficulty in rejecting unseen, related lure words in the RH, consistent with a broader spread of activation of semantic associates.

In a variation on the methodology, Fabiani, Stadler, and Wessels (2000) presented DRM word lists to the LVF or RVF during training, thus participants encoded word pairs in either the LH or RH. During subsequent retrieval, target words were presented centrally. They found that response times to lure words that corresponded to word lists presented initially to the RH were longer than lure words for lists presented to the LH. Thus, words encoded in the RH elicited a broader spread of activation, meaning participants took longer to retrieve the information needed to accurately accept or reject the words. However, no difference in performance accuracy between the LH and RH was reported, which would be expected if the RH elicited a broader spread of activation. One explanation for this could be that words were presented twice during the learning phase, and thus participants may have been able to process this information across hemispheres.

Lateralisation studies investigating the effects of LH and RH semantic processing on false memory performance have also aimed to investigate conditions in which false memories are more or less likely to occur in either hemisphere. Faust, Ben-Artzi, and Harel (2008) investigated the effect of ambiguity of lure words on false memory performance for word lists that were related to the dominant or subordinate meaning of the lure words. They found that more lure words related to dominant lists were falsely recognised in the LH than the RH, whereas more ambiguous lure words semantically related to the subordinate lists were falsely recognised in the RH, further supporting the suggestion that the RH has a broader spread of activation of semantic associates.

1.5.3 Sleep’s Role in Hemispheric Processing of Memories

There is evidence to suggest that the RH is more dominant than the LH during a period of sleep (Casagrande & Bertini, 2008a; Casagrande, Violani, De Gennaro, Braibanti, & Bertini, 1995). Whereas the LH is more active during wakefulness, the RH has been found to be more active during sleep (Bolduc, Daoust, Limoges, Braun, & Godbout, 2003).There are two main theories proposed to explain why there is this superiority of the RH during sleep. The first theory suggests that the RH is more vigilant overnight (Whitehead, 1991), whereas the second theory, the homeostatic hypothesis, suggests that the LH, which is much more active during wake, needs time to restore during sleep to a much greater extent than the RH and thus falls asleep first, and takes over from RH processing once it is restored (Kattler, Dijk, & Borbely, 1994).

Casagrande and Bertini (2008b) aimed to test these two hypotheses by giving participants a finger-tapping task during different stages of wakefulness, including sleep onset, awakenings from NREM sleep in the first half of the night, and awakenings from REM sleep in the second half of the night. They found greater activation of the LH during wake, whereas there was a shift in hemispheric activation to the RH when the task was administered to participants after being awakened from both NREM and REM sleep. Since there was a RH dominance in both the first and second half of the night, these findings support a RH advantage for vigilance during sleep, rather than the homeostatic hypothesis.

In contrast, Violani, Testa, and Casagrande (1998) found evidence to support the homeostatic hypothesis. They found that the right hand, and therefore the LH, showed greater motor activity during wake, and that this activation switched to the RH during the first half of a night of sleep. However, during the second half of a night of sleep the LH showed greater activation. Thus, the LH was able to restore during the first half of sleep, and so became more active in the second half of the night of sleep, supporting the homeostatic hypothesis.

1.5.4 Sleep's Role in Hemispheric Processing of False Memories

It is therefore evident that there is a RH advantage at least during the first half of a night of sleep, and a LH advantage during wakefulness. The increase in lure word acceptance found after sleep (Diekelmann et al., 2010; Payne et al., 2009) could therefore be due to the RH broad semantic network leading to a broader spread of activation during sleep. If this is the case, then sleep should increase RH false memories to a greater extent than LH false memories. Monaghan et al., (2017) tested this hypothesis, investigating the effect of sleep on the hemispheric processing of veridical and false memories. The study supported previous findings of an increase in false memory performance after a period of sleep compared to wakefulness (Diekelmann et al., 2010; Payne et al., 2009). However, there was no significant difference in false memory performance between the LH and RH, suggesting that sleep therefore supports spreading activation equally across the hemispheres, as opposed to enhancing the RH broad semantic network. It was also found that there was a general shift in processing of all word types to the LH after a period of sleep, supporting previous research to suggest a LH advantage for word recognition testing (Pirozzolo & Rayner, 1977).

Conversely, a similar study found that in a nap paradigm there was a RH bias for false memories, and this correlated with RH sleep spindle density (Shaw & Monaghan, 2017). This difference between the two studies might be explained by a homeostatic hypothesis, such that the RH is more vigilant in the first half of a night of sleep, whereas the LH is more active in the second night-half. As such, in a nap paradigm the RH is more likely to remain vigilant whilst the LH restores, providing an explanation for an increase in RH false memories correlating with RH sleep spindles (Shaw & Monaghan, 2017). However overnight, as the LH becomes more active, the

RH bias is weakened and equal spread of activation across hemispheres is likely to occur (Monaghan et al., 2017). This explanation is yet to be explicitly examined, thus it remains unclear whether sleep enhances the bias for RH false memories or supports equal spread of activation. Further research is therefore needed to understand the exact mechanisms underlying this potential influence.

1.6 Emotional Memory Consolidation

1.6.1 Effects of Emotion on Veridical Memory Consolidation

Not all information is remembered equally, and research investigating emotional memory performance has indicated that memories for emotional events and information is recalled much more accurately than neutral memories (van der Helm & Walker, 2009; Walker, 2009). It is suggested that this enhancement for emotional memories is due to co-activation of the amygdala and hippocampus (Walker, 2009), with increased activity in the amygdala correlating with increased emotional memory (Cahill et al., 1996; McGaugh, 2004). However, studies of emotional memory focus on two separate dimensions of memory: valence and arousal. LaBar and Cabeza (2006, p. 54) categorise valence as “a dimension of emotion that varies from unpleasant (negative) to pleasant (positive), with neutral often considered an intermediate value”, whereas arousal is categorised as “a dimension of emotion that varies from calm to excitement”.

Previous literature suggests that both emotional arousal and emotional valence are associated with improvements in memory performance. Kensinger and Corkin (2003) examined the independent effects of negative versus positive valence and high and low arousal on subsequent memory performance. They found that words of negative valence, and those of high arousal, led to increased recognition compared to neutral or low arousal words, with words high in arousal leading to the greatest influence on memory performance. Similarly, Bradley, Greenwald, Petry, and Lang (1992) found a significant increase in memory performance for stimuli that was rated as high in arousal, and Lang, Dhillon, and Dong (1995) found that television messages of high arousal, regardless of valence, were better remembered than low arousal messages. When arousal was controlled, positive television messages were remembered better than negative messages. However, there are some variations found within the literature as to the extent to which emotional valence and arousal influence memory performance. Adelman and Estes (2013) investigated the influence of both emotional valence and emotional arousal of words, and found that words of both negative and positive valence were remembered significantly more than neutral words regardless of arousal.

Figurative Language

One type of stimuli high in emotionality is that of figurative language. Figurative language is often rated as more emotional than literal counterparts (Citron & Goldberg, 2014; Citron, Güsten, Michaelis, & Goldberg, 2016), and such expressions are used extensively in everyday communication (*e.g., fall in love, a rough day*) (Carter, 2004; Jackendoff, 1995; Pollio, Barlow, Fine, & Pollio, 1977). The Conceptual Metaphor Theory (Lakoff & Johnson, 1980) proposes that emotional experiences are often difficult to describe using literal concepts, and thus metaphorical language is used as a way of defining and conceptualising abstract emotional experiences in a more concrete way. In support of this theory, Fainsilber and Ortoney (1987) found that metaphorical language was used more often when describing subjective feelings than concrete actions. Furthermore, subjective feelings that were higher in intensity were described using metaphorical expressions to a greater extent than less intense emotions.

Neuroimaging studies investigating brain responses to figurative language have found activation of the amygdala and hippocampus when processing metaphorical expressions (Citron & Goldberg, 2014; Citron et al., 2016). Increased activation of the amygdala and co-activation of the hippocampus and the amygdala has been linked to increases in the processing of other types of emotional stimuli (Cahill et al., 1996; McGaugh, 2004; Walker, 2009), thus supporting a role of figurative language for conceptualising emotional experiences.

1.6.2 Effects of Emotion on False Memories

Emotion has also been found to influence false memories, however whether emotion reduces or increases the false memory effect is debated. Pesta, Murphy, and Sanders (2001), found that emotional lure words were falsely remembered significantly less than non-emotional lures. However, emotional words were only presented to participants at testing, and thus these words were much more distinctive to participants and so easily recognised and rejected. Thus, in a second experiment, Pesta et al. (2001) added three emotional words to the to-be-remembered lists during training. The inclusion of emotional words during encoding reduced their distinctiveness, increasing the false memory effect for emotional lure words, although non-emotional lure words were still falsely remembered at a higher rate than emotional lure words. These findings are supported by Kensinger and Corkin (2004), who found that distinctiveness reduced emotional false memories in both younger and older adults.

When distinctiveness is controlled for, there is evidence for an increase in false memories for emotional information, but only in tests of recognition. False recognition of negative unseen, related lure words was found to be higher than false recognition of neutral words, however this did not extend to tests of false recall (El Sharkawy et al., 2008; Howe et al., 2010). Similarly, Brainerd, Stein, Silveira, Rohenkohl, and Reyna (2008) found that negative word lists led to increased false recognition of critical lures, whereas positive word lists led to a decrease in false recognition in comparison to neutral words. However, research conducted by Bauer, Olheiser, Altarriba, and Landi (2009) did find an effect of emotion on false recall of words. Thus, effects of emotion on false recognition are consistent throughout the literature, whereas effects on false recall are still under review.

1.6.3 The Role of Sleep in Emotional Memory Consolidation

There are suggestions that sleep plays a selective role in the consolidation of memories, such that memories are filtered for those that will be selectively reactivated during sleep. In particular, sleep has been found to preferentially enhance memory for both negative and positive emotional information compared to non-emotional (van der Helm & Walker, 2009; Walker, 2009).

REM sleep in particular has been observed to support the strengthening of emotional memories. Increasing levels of limbic and forebrain acetylcholine has been found to be a specific characteristic of REM sleep (Vazquez & Baghdoyan, 2001), and is suggested to play a significant role in improving emotional memory performance (McGaugh, 2004). Studies from the night-half paradigm indicate that late sleep rich in REM supports emotional memory, whereas early SWS-rich sleep has no effect on emotional memory (Groch, Wilhelm, Diekelmann, & Born, 2012; Wagner, Gais, &

Born, 2001). Furthermore, Nishida, Pearsall, Buckner, and Walker (2009) showed that those who napped after encoding images showed greater consolidation of emotional than neutral stimuli, and this correlated with increases in REM sleep.

The proposed REM sleep advantage on memory performance may be due to a broader spread of activation of emotional compared to neutral stimuli. Carr and Nielsen (2015) found increased priming of emotional words after a nap containing increased REM sleep, thus REM sleep led to increased integration and consolidation of emotional memories, but also increased the spread of activation of semantic networks to other primed associates. This theory raises questions as to whether REM sleep may also enhance false memories for emotional stimuli either by activating a broader set of associations with the to-be-remembered stimuli or increasing gist-based memory representations.

Despite extensive research indicating an effect of sleep on the consolidation of emotional veridical memories, only one study has attempted to examine the influence of sleep on the formation of emotional false memories. McKeon, Pace-Schott, and Spencer (2012) using negative and neutral DRM word lists, found that a period of overnight sleep increased veridical recall for neutral word lists, but not for negative word lists. This finding goes against research to suggest that sleep improves emotional memory performance compared to neutral (Chambers & Payne, 2014; Hu, Stylos-Allan, & Walker, 2006; Nishida et al., 2009; Payne et al., 2008). Furthermore, McKeon et al. found an increase in both false recall of neutral and negative lure words after sleep, supporting previous research that indicates that sleep leads to a broader spread of activation of memories, or memory for the general gist, but does not support a differential effect of sleep on memory for emotional versus non-emotional information. However, this study employed a test of recall to examine false memories, and thus we may expect to see differences between emotional and non-emotional false memories in tests of recognition. Indeed, studies of false memories without sleep have found an effect of emotion only in tests of recognition (El Sharkawy et al., 2008; Howe et al., 2010).

Whilst the majority of research suggests a role of REM sleep in the consolidation of emotional memories, NREM sleep spindles have also been associated with increases in emotional memory processing. By pharmacologically increasing the number of sleep spindles, Kaestner, Wixted, and Mednick (2013) found increased memory performance to high arousal stimuli, in particular those of negative valence. Furthermore, slower response times to neutral compared to emotional memories has been correlated with increasing sleep spindles, suggesting a suppression of neutral information that indirectly influences emotional memories (Cairney, Durrant, Jackson, & Lewis, 2014).

1.6.4 Sleep to Forget, Sleep to Remember

Not only has sleep been found to influence how well we consolidate emotional memories, it is also proposed to influence the affective experience of the emotion. As such, the “sleep to forget, sleep to remember” (SFSR) hypothesis proposes that whilst REM sleep strengthens memory for emotional information, it also reduces the emotional tone of the memory (van der Helm & Walker, 2009) (see Figure 3). In support of the SFSR hypothesis, van der Helm, Yao, Dutt, Rao, Saletin, and Walker (2011) found that sleep decreased amygdala reactivity in response to previously seen emotional images, and this decrease related to a reduction in subjective ratings of emotion. Furthermore, REM sleep in particular was associated with decreased amygdala reactivity and reduced emotional ratings.

However, this is only one of a few studies that support a SFSR hypothesis. Further investigations show that overnight sleep does not reduce the emotional experience of information (Wiesner et al; 2015) and Wagner, Fischer, and Born (2002) found that sleep rich in REM actually led to increased negative ratings of aversive images, whereas SWS-rich sleep increased positive ratings, and overall sleep enhanced ratings of arousal. A potential explanation for this lack of an effect of sleep on reducing the affective tone of the stimuli is that the affective tone may be reduced over a number of nights of sleep, and thus one night has little or no effect. Further research is therefore needed to understand whether the affective tone of emotional memories can be reduced after one night of sleep, or whether a number of nights of sleep is required.

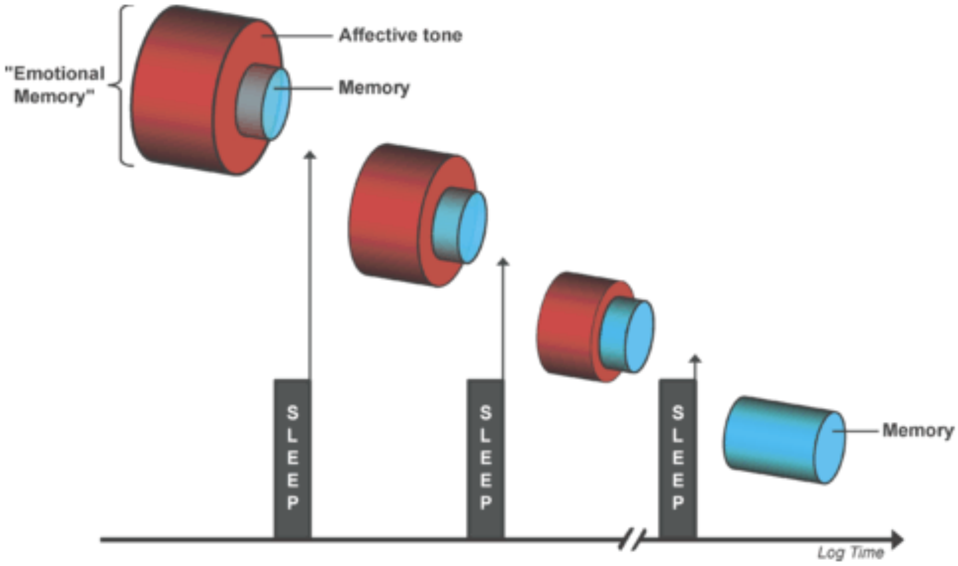


Figure 1.3. Model of sleep dependent emotional memory processing: A sleep to forget and sleep to remember hypothesis. Source: (Walker, 2009)

1.7 Research Objectives of the Thesis

The aim of the thesis is to investigate the role of sleep on both veridical and false memories, and to explore factors that influence which memories are preferentially consolidated during sleep. The literature thus far suggests a beneficial role of sleep in declarative memory, however the role of sleep in the formation of false memories is unclear. Furthermore, research investigating the specific influences of lateralisation effects and emotionality on both veridical and false memories is under-represented. This thesis seeks to address some of these questions and gain a greater understanding of factors that influence sleep-dependent memory consolidation.

More specifically, Chapter 2 aimed to provide a meta-analytical review of the literature on sleep and DRM memories thus far to establish the strength of the effect of sleep on DRM memories, and how various methodological differences may influence the effect of sleep on memory performance. In Chapter 3, participants were presented with DRM word lists to the LH or RH during encoding, to investigate whether the processing of memories presented to the LH or RH influences subsequent memory performance, and to expand on previous literature of lateralisation effects after sleep (Monaghan et al., 2017, Shaw & Monaghan, 2017). In Chapter 4, the focus shifts towards emotion, firstly to investigate whether sleep's selective enhancement of veridical emotional memories can be extended to false memories. Although current findings indicate no specific role of sleep in false memories for emotional stimuli (McKeon et al., 2012), differential effects of negative and positive emotion are yet to be fully explored. Thus, we investigated whether variations in valence, when controlling for arousal, significantly impacted on both veridical and false memories, and whether sleep interacted with emotionality to support memory consolidation. In Chapter 5 this study was extended, using a nap paradigm and PSG to record participants' sleep architecture so as to relate sleep stages to emotional memory consolidation. In Chapter 6, the generality of emotional memory effects was extended beyond the DRM paradigm, by testing the effect of sleep on the learning of novel figurative language. Since metaphorical expressions are often rated as highly emotional, I investigated the role of sleep on memory for negative, positive, and neutral metaphorical word pairs. I also gained affective ratings in order to examine whether the affective tone of such emotional expressions can be reduced overnight. Finally, chapter 7 discusses the key findings throughout the thesis, explores future directions, and outlines methodological issues of the current studies.

The studies in this thesis will therefore contribute to current models of memory consolidation, and theories of sleep's role in memory consolidation. If lateralisation effects on memory are extended to encoding, then theories of hemispheric processing of memory consolidation will need to be updated to reflect the role of encoding processes on memory. Similarly, investigating the role of emotion on memory will allow for a greater understanding of the effects of valence and arousal on both specific and gist-based memory consolidation. Furthermore, models of sleep's role in memory consolidation may need to be updated to reflect the differential effects of valence and arousal, and the influence of sleep stages on emotional veridical and false memories.

Chapter 2: When Does Sleep Affect Veridical and False Memory Consolidation? A Meta-Analysis.

The review of the literature thus far indicates that the effect of sleep on veridical and false memories with the DRM paradigm may not be robust across studies, but instead may be influenced by a number of moderating variables. One clear inconsistency throughout the literature is that of recall versus recognition testing, however the literature review also outlined other potential factors that influence DRM memories. We therefore conducted a meta-analytic review of studies investigating DRM memory performance after sleep versus wakefulness, and explored the influence of recall versus recognition testing, as well as other potential moderators, on sleep-dependent memory consolidation.

Statement of Author Contribution

In the chapter entitled "When Does Sleep Affect Veridical and False Memory Consolidation? A Meta-Analysis", the authors agree to the following contributions:

Chloe R. Newbury – 80% (Design, data collection, data analysis, and writing)

Signed: C. Newbury Date: 25/04/2019

Professor Padraic Monaghan – 20% (Design and review)

Signed: Date: 25/04/2019

Text as it appears in Newbury, C. R., & Monaghan, P. (2018). When does sleep affect veridical and false memory consolidation? A meta-analysis. *Psychonomic Bulletin & Review*, 1-14. doi:10.3758/s13423-018-1528-4.

Abstract

It is widely accepted that sleep aids in the encoding, consolidation, and retrieval processes involved in memory processing, however, the conditions under which sleep influences memory may be substantially constrained. In a meta-analysis, we examined the effect that sleep has on both veridical (accurate) and false memory consolidation, in studies using the Deese-Roediger-McDermott (DRM) paradigm for memory of thematically-related words. The meta-analysis revealed that, whereas there was no overall effect of sleep on either accurate or false memories, the effect of sleep on memories was moderated by two constraints. First, sleep effects were influenced by the number of words within each themed word list, relating to differences in processing the associative network of related words. Second, sleep effects were greater in recall than recognition tests. Thus, whether sleep consolidation increased or decreased DRM veridical or false memory effects depended on specific features of the memory task.

1. Introduction

Sleep benefits both the encoding and retrieval processes involved in memory consolidation, improving both declarative and procedural memory compared to the same time spent awake (Rasch & Born, 2007; Rasch & Born, 2013; Stickgold, 2005; Walker & Stickgold, 2006). The Active Systems Consolidation hypothesis (Diekelmann & Born, 2010; Marshall & Born, 2007; Rasch & Born, 2013), suggests that information and events that we are exposed to during wakefulness are encoded initially in the hippocampus and neocortical systems. Consolidation during sleep then leads to repeated reactivation of these encoded memory representations, leading to an integration of selective information to the neocortex, where the memory is established in the long-term store (Lewis & Durrant, 2011). Substantial evidence supports this theory, for example declarative memory for word pairs has been found to be greater after a delay including a period of sleep compared to wake (Gais & Born, 2004; Plihal & Born, 1997; Wilson, Baran, Pace-Schott, Ivry, & Spencer, 2012).

Several studies have also tested the hypothesis that sleep not only affects processing and consolidation of previously experienced material, but also impacts the formation of false memories. The Deese-Roediger-McDermott (DRM) paradigm (Roediger & McDermott, 1995) has been extensively used to test when unseen, related information, termed false memories, is activated in memory. In this paradigm, participants are exposed to lists of semantically related words (eg., *bed, dream, tired, snooze, yawn,* etc.), and are asked to recall or recognise words previously seen in the initial lists. Words are categorised as either those that appeared in the initial lists (old words), words that did not appear in the lists but were closely related, known as lure words (eg., *sleep* in the above list), or unseen, unrelated words (new words). Participants are more likely to recall, or identify as previously seen, lure words than new words, demonstrating the false memory effect (McDermott, 1996; Roediger & McDermott, 1995; Roediger, Watson, McDermott, & Gallo, 2001).

Whilst evidence for sleep's effect on veridical memory performance has been widely replicated, the question of whether sleep has an effect on DRM false memories remains. Potential inconsistencies in results emerge between tests of recall, where false memories seem to be enhanced by sleep (Diekelmann, Born, & Wagner, 2010; Payne et al., 2009), and tests of recognition, where sleep has been observed to either enhance, have no effect, or even reduce false memories (Diekelmann, Landolt, Lahl, Born, & Wagner, 2008; Fenn, Gallo, Margoliash, Roediger, & Nusbaum, 2009; Monaghan, Shaw, Ashworth-Lord, & Newbury, 2017). The activation/monitoring framework (Collins & Loftus, 1975) provides one possible explanation for the differences found between DRM tests of recall and recognition. The framework proposes that during tests of recognition, monitoring cues are activated when the words are presented to participants, allowing for the suppression of related but unseen words (Watson, McDermott, & Balota, 2004). During tests of recall, these monitoring cues are not available, and so a greater number of associated words are activated. This leads to greater false memory in tests of recall than recognition. Sleep has been found to improve source-monitoring abilities (Johnson, Hashtroudi, & Lindsay, 1993), and therefore improves the ability to reject unseen, related items during tests of recognition to a greater extent than during recall.

This difference in memory performance between tests of recall and recognition has been suggested in a meta-analysis of only a small number of studies that were published at the time (Chatburn, Lushington, & Kohler, 2014). A small non-significant effect of sleep on false recognition was found, whereas false recall led to a large significant increase in false memory. However, this study examined the overall effect of only four studies in total, two studies on false recognition, and two on false recall. Therefore, the reliability of the effect of sleep on both false recall and recognition is still under review. To address this, the current meta-analysis includes a larger sample of DRM studies, with five individual experiments examining the effect of sleep on false recall, and eight experiments investigating false recognition. This allows for a more detailed exploration as to the effects of the two methods of testing, and a greater understanding as to whether the effect of sleep on false memories does in fact reliably differ between tests of recall and recognition.

The larger number of experiments analysed in this meta-analysis also permits investigation of other potential moderator variables that may contribute to the effect of sleep on memory consolidation and the production of false memories within the DRM paradigm. In particular, we can determine whether the number of words in each list and the total number of lists that participants are required to remember influences the false memory effect. Using the DRM paradigm in a standard memory test (so not testing the effect of sleep), Robinson and Roediger (1997) investigated the effect of varying list lengths on false recall and recognition. They found that increasing list length led to increases in both false recall and false recognition. Robinson and Roediger (1997) suggested that a larger number of words in each list increases the opportunity for participants to develop associations between the words, and therefore primes a larger number of unseen, related words during testing.

A possible explanation for generation of false memories in DRM tests is that of spreading activation (Collins & Loftus, 1975). Word lists that participants are exposed to can activate unseen words that are similar in meaning to previously seen words. The Associative Activation Theory (AAT) of false memories suggests that these lure words are activated due to their similarity or association with the seen words (Howe, Wimmer, Gagnon, & Plumpton, 2009; Roediger et al., 2001). Lists with greater strength of semantic association with the critical lure elicit increased false memories than those with weaker associations, due to spreading activation among associates within semantic memory (Gallo & Roediger, 2002). Alternatively, Fuzzy Trace Theory (FTT; Payne et al., 2009) argues that false memories are a consequence of participants determining the gist or general theme of a list, and then activating all words related to that general meaning (Howe & Wilkinson, 2011). The mechanism of gist generation could again be due to spreading activation, with the theme generated as a consequence of interactive activation among associated words.

If sleep leads to greater spread of activation of previously seen word lists due to AAT or FTT (as proposed by Cai, Mednick, Harrison, Kanady, Mednick & Albright, 2009; Sio, Monaghan & Ormerod, 2013), then we should expect to see an increase in false recall and recognition of lure words after sleep in comparison to wakefulness. These theories raise predictions about the extent to which manipulating the density of inter-relations between words in a thematically-related list affects the role of sleep in consolidation. A longer list of related words is more densely interconnected (Robinson & Roediger, 1997), and so spreading activation will occur to a greater extent for both sleep and wake groups equally. Thus, the benefit of sleep-related spreading activation is less likely to be detected than in a shorter list of related words, where the lure word concept receives only weak activation from a small set of related words within semantic associative memory (see Shaw & Monaghan, 2017, for a similar argument related to hemispheric processing). Hence, list length may be a critical factor in determining whether veridical and false memories are promoted by sleep. Indeed, previous research suggests that sleep is more beneficial when task difficulty increases, for both motor skills tasks (Stickgold & Walker, 2004), and problem-solving tasks (Sio et al., 2013). If increasing the number of words in each list leads to closer associations and so easier access to semantically similar lure words, then we would expect sleep to increase false memories for studies with fewer words in each list, as activation of the lure word, or the theme, is more difficult to accomplish, so greater spreading activation is required across semantic networks.

Similar principles could also be expected to apply to the number of different lists that participants are exposed to. For instance, source monitoring is likely to become more difficult with larger numbers of lists, thereby increasing the likelihood of false memories, and decreasing veridical memory. Spreading activation across a large number of distinct thematic lists may also mitigate the potential effect of sleep on the generation of false memories.

The modality of presentation of word lists has also been found to affect the formation of DRM false memories. Previous research suggests differences in performance following visual compared to auditory presentation of word lists. For both tests of recall and recognition, research indicates a significant reduction in the false memory effect when words are presented visually (Kellogg, 2001; Smith & Hunt, 1998). However, this difference in the effect of modality on false memory performance has been found to only be significant in those participants with higher working memory capacity (Smith & Engle, 2011). This difference in performance between visually and auditorily presented word lists was not found for veridical memory (Smith & Engle, 2011; Smith & Hunt, 1998). It is therefore of interest to assess modality as a potential moderator in the current meta-analysis.

The emotionality of to-be-remembered word lists may also influence the size of the effect of sleep on both accurate and false memories. Research indicates an increase in overall memory performance for information with positive or negative emotional valence (Adelmann & Estes, 2013; Kensinger & Corkin, 2003). Further, emotionality of word lists has also been found to increase false recognition in DRM tests (Howe, Candel, Otgaar, Malone & Wimmer, 2010; Sharkawy, Groth, Vetter, Beraldi & Fast, 2008), however possible differences arise between lists of negative and positive valence, with an increase in false recognition of negative word lists, and a decrease in false recognition of positive lists, compared to lists rated as neutral (Brainerd, Stein, Silveira, Rohenkohl & Reyna, 2008). The effect of emotionality on false recall is less clear. Bauer, Olheiser, Altarriba and Landi (2009) suggest an increase in false recall for emotional word lists, whereas Howe et al., (2010) suggest a reduction in false recall for emotional compared to neutral word lists.

Sleep is suggested to further enhance this bias for the consolidation of emotional information, with studies indicating a role of rapid eye movement (REM) sleep specifically in the processing of emotional memories (Carr & Nielsen, 2015; Goldstein & Walker, 2014). Cai et al., (2009), and Carr and Nielsen (2015) suggest that REM sleep increases spreading activation, hence performance differences may be evident between emotional and neutral word lists after sleep. We thus tested emotionality of word lists as a potential moderator in the current meta-analysis, to assess whether emotionality leads to an enhanced effect of sleep compared to wake on both veridical and false memories.

This larger set of studies included in a meta-analysis of DRM sleep-related effects means we can also assess daytime nap versus overnight sleep effects on veridical and false memories. If sleep leads to greater spreading activation to semantic associates (Collins & Loftus, 1975), then we would expect an increase in time spent asleep to result in improved veridical performance as well as enhanced false memories. Since Cai et al., (2009) suggest that REM sleep, which occurs to a greater degree in the latter half of a night's sleep, increases spreading activation, performance differences between sleep and wake groups may be more significant with overnight sleep than a short nap. Furthermore, Payne et al. (2009) found a negative correlation between veridical recall and slow wave sleep (SWS), indicating reduced veridical recall performance with increasing SWS, again suggesting that differences between the sleep and wake groups may be more significant during a longer period of sleep compared to a daytime nap.

In this current meta-analysis, we therefore aimed to analyse what effect sleep has on both accurate and false memory in DRM tests. We included six potential moderator variables, and analysed the possible effect that these may have as constraints on effects of sleep on memory consolidation: (1) whether the memory task is recall or recognition testing, (2) the number of words in each list, (3) the number of different lists learned, (4) whether words were presented auditorily or visually, (5) emotionality of the lists, and (6) whether the study was an overnight or nap study.

Analysing sleep effects on old, new, and lure words individually is useful for formulating comparisons between recall and recognition tests. However, in recognition tests, any observed changes in accuracy as a consequence of sleep could be due to changes in discriminability between word types or changes in response biases to respond yes more or less often. We therefore also used Signal Detection measures to distinguish the overall sensitivity or discriminability (d') and response bias (C) between sleep and wake groups for the studies testing recognition memory. We distinguished true recognition, defined as differences in responses to old words and new (unrelated) words, and false recognition, defined as differences in responses to lure words and new words. We hypothesised that sleep groups would have a larger discriminability and response bias score than wake groups for true recognition, which would indicate that the sleep groups are more likely to correctly accept old words as previously seen, and accurately reject new words as unseen. If so, then this would provide evidence in support of a positive role of sleep on memory consolidation and improving accuracy of memory. In contrast, the effects of sleep on false recognition are still under review, and so we might expect to see a larger discriminability and response bias score for the sleep groups if sleep increases false recognition (Monaghan et al., 2017), larger scores for the wake groups if sleep reduces false recognition (Fenn et al., 2009), or no difference in discriminability and response bias if sleep does not influence false recognition (Diekelmann et al., 2008).

2. Methods

To collect the relevant data, we conducted searches in both Scopus and Web of Science [23-06-2017], using the keywords “sleep OR nap AND false memories”. Scopus produced 113 results, and Web of Science produced 139 results. Our next step was then to check for duplicates, yielding a total of 169 unique entries. An additional 2 papers from our own research lab were also included in the final analysis, although these were not produced during the main searches due to being submitted for review or in preparation at the time of the searches. These entries were then screened using the following inclusion criteria: (1) behavioural studies conducted with adult participants, who were (2) exposed to DRM word lists and (3) asked to take part in a recall or recognition task (4) after a period of sleep (which could be overnight or a nap), with (5) a wake group comparison condition. This screening led to the inclusion of nine papers in total, with some papers containing multiple experiments (13 individual experiments with a total of 596 participants overall; see Table 1 for summary data and moderators for each experiment).

2.1 Meta-Analysis

The effect sizes reported are the standardized mean difference in proportion of responses to each word type given as old (in the recognition tests) and proportion recalled (in the recall tasks) between the sleep and wake group, with positive values meaning increased proportion of responses in the sleep group than the wake group. Effect sizes for previously seen (old) words, unseen, related (lure) words and unseen, unrelated (new) words were calculated and analysed separately (see Table 1 for means and effect sizes). True recognition and false recognition d' and C sensitivity measures were also computed for the studies testing recognition memory. When not enough data was provided in the paper to calculate effect size and sensitivity measures, authors were contacted for means and standard deviations.

We computed Hedge's *g* based on the means and variance reported in each study for the wake and sleep groups. Hedge's *g* is a variation of Cohen's *d* that corrects for biases due to small sample sizes. We then fitted a random effects model using the R package metafor (Viechtbauer, 2010). A random effects analysis was chosen because this method, in contrast to a fixed effects meta-analysis, allows for inconsistencies between the studies analysed, calculation of possible sampling error, and assessment of the effects of moderators on the size of the effect (Borenstein, Hedges, Higgins, & Rothstein, 2010). We introduced the six moderator variables, (1) recall or recognition testing, (2) number of words in each list, (3) number of lists learned, (4) whether words were presented auditorily or visually, (5) emotionality of lists, and (6) overnight or nap study to the model, to examine any possible influence of these moderators on the effect size of sleep.

Table 2.1

Descriptions of means, standard deviations, Hedge's g and, standard error for lure, old, and new words for each experiment, as well as information on moderators.

		Lure Words				Old Words				New Words									
		Sleep	Wake			Sleep	Wake			Sleep	Wake				Number of Words in		Presentation		Sleep
Authors	Year	Mean (SD)	Mean (SD)	Hedge's g	SE	Mean (SD)	Mean (SD)	Hedge's g	SE	Mean (SD)	Mean (SD)	Hedge's g	SE	Test Type	Lists	List	Type	Emotionality	Type
Payne et al. (Experiment 1)	2009	3.600 (1.649)	2.900 (1.697)	0.42	0.03	21.900 (9.895)	15.700 (7.637)	0.70	0.03	5.600 (7.422)	6.200 (5.940)	-0.09	0.03	recall	8	12	auditory	Neutral	overnight
Payne et al. (Experiment 3)	2009	4.300 (1.600)	2.900 (1.497)	0.88	0.14	26.600 (12.000)	26.400 (15.715)	0.01	0.13	8.090 (5.200)	8.600 (9.354)	-0.08	0.13	recall	8	12	auditory	Neutral	nap
Diekelmann et al. (low performers)	2010	2.860 (1.323)	1.000 (1.323)	1.32	0.31	25.140 (24.780)	21.430 (9.366)	0.19	0.25	4.710 (2.884)	4.430 (2.884)	0.09	0.25	recall	8	12	auditory	Neutral	overnight
Diekelmann et al. (high performers)	2010	1.640 (1.194)	2.270 (1.194)	-0.51	0.17	43.730 (31.130)	50.360 (9.386)	-0.28	0.17	2.820 (2.885)	4.180 (2.885)	-0.45	0.17	recall	8	12	auditory	Neutral	overnight
McKeon et al.	2012	2.267 (1.579)	0.800 (0.862)	1.12	0.15	14.067 (5.675)	7.600 (4.733)	1.20	0.15	2.133 (1.685)	2.667 (2.845)	-0.22	0.13	recall	10	10	auditory	Emotional	overnight
Fenn et al. (Experiment 1)	2009	0.690 (0.203)	0.760 (0.203)	-0.34	0.04	0.560 (0.920)	0.590 (0.136)	-0.05	0.04	0.180 (0.136)	0.220 (0.203)	-0.23	0.04	recognition	16	15	auditory	Neutral	overnight
Fenn et al. (Experiment 2)	2009	0.740 (0.120)	0.860 (0.120)	-0.97	0.13	0.720 (0.640)	0.710 (0.120)	0.02	0.12	0.150 (0.160)	0.190 (0.160)	-0.24	0.12	recognition	10	15	visually	Neutral	overnight
Fenn et al. (Experiment 3)	2009	0.630 (0.170)	0.730 (0.167)	-0.59	0.06	0.570 (0.640)	0.620 (0.167)	-0.10	0.06	0.210 (0.113)	0.310 (0.167)	-0.69	0.07	recognition	10	15	visually	Neutral	overnight
Diekelmann et al. (Experiment 1)	2008	0.770 (0.116)	0.750 (0.112)	0.17	0.13	0.670 (0.300)	0.700 (0.112)	-0.13	0.13	0.270 (0.116)	0.280 (0.112)	-0.09	0.13	recognition	18	15	auditory	Neutral	overnight
Lo et al.	2014	0.730 (0.187)	0.870 (0.112)	-0.88	0.15	0.740 (0.150)	0.780 (0.075)	-0.33	0.14	0.510 (0.224)	0.520 (0.224)	-0.04	0.13	recognition	10	15	auditory	Neutral	overnight
Monaghan et al.	2017	0.726 (0.159)	0.574 (0.153)	0.96	0.08	0.668 (0.167)	0.606 (0.134)	0.40	0.07	0.221 (0.170)	0.215 (0.178)	0.04	0.07	recognition	12	10	visually	Neutral	overnight
Newbury & Monaghan	submitted	0.587 (0.125)	0.625 (0.178)	-0.25	0.07	0.520 (0.132)	0.561 (0.158)	-0.28	0.07	0.237 (0.160)	0.222 (0.167)	0.09	0.07	recognition	15	12	visually	Emotional	overnight
Shaw & Monaghan	2017	0.740 (0.172)	0.620 (0.172)	0.68	0.13	0.649 (0.124)	0.568 (0.124)	0.64	0.13	0.177 (0.204)	0.271 (0.204)	-0.45	0.12	recognition	12	10	visually	Neutral	nap

3. Results

3.1 Lure words

The overall effect size for the mean difference in the proportion of responses to lure words given as old between the sleep and wake group, measured by Hedge's *g*, was 0.129 (SE = .210), which indicated no significant difference from zero (95% CI [-0.282, 0.539], $p = .540$). See Figure 1 for a forest plot of effect sizes. Since at the time of data analysis one study within the meta-analysis was unpublished (Newbury & Monaghan, submitted), we conducted a second analysis without this dataset. The overall effect size did not significantly change, (Hedge's $g = 0.165$, SE = 0.227, 95% CI [-0.281, 0.610], $p = .469$), and so we continued our analysis of the full dataset. There are possible moderator variables that may lead to differing directions of effects, highlighted by the significant heterogeneity ($Q(12) = 63.227$, $p < .001$), indicating that there is variance in the data that cannot be explained by the random measurement error. We therefore analysed the effects of each of the moderators (see Table 2 for significance of each moderator).

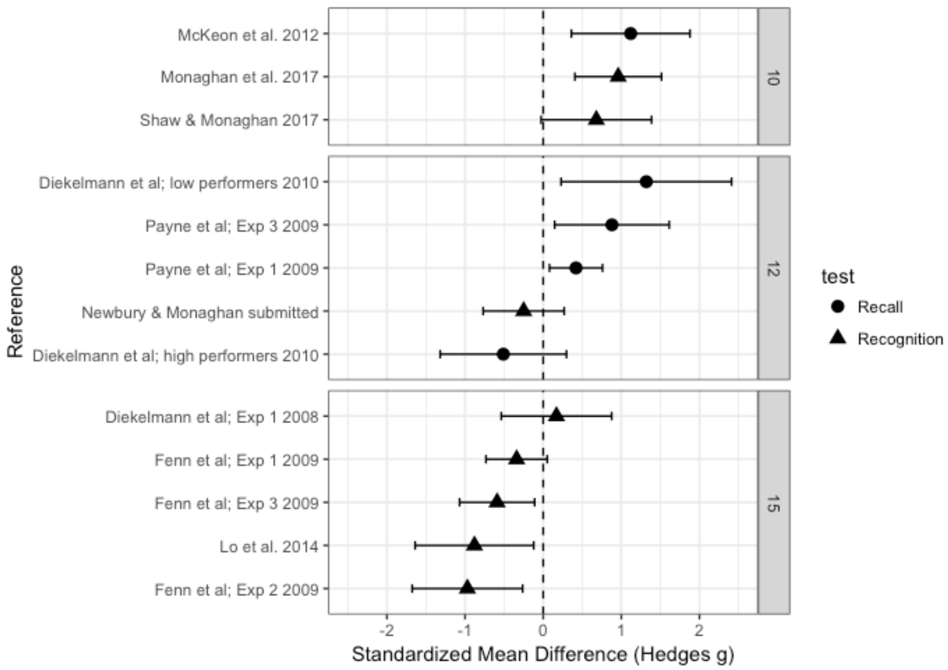


Figure 2.1. A forest plot containing effect sizes and 95% confidence intervals for the difference in proportion of old responses between sleep and wake groups for lure words. Studies are split by two moderators: number of words in DRM lists, and recall versus recognition studies. Effect sizes further to the right indicate more lure words falsely recalled or recognised as old in the sleep group than the wake group, and therefore increased false memories after sleep.

Table 2.2

The effect of each moderator on the overall effect size difference between sleep and wake groups for lure words.

Moderator	*df*	Heterogeneity (Q)	*p*
Recall versus Recognition	1	3.685	.055+
Number of lists	1	0.291	.590
Number of words in each list	2	18.368	<.001***
Auditory versus Visual	1	0.387	.534
Emotional versus Neutral	1	0.264	.608
Nap versus Overnight sleep	1	1.818	.178

Note: +*p* < .1, ****p* < .001

3.1.1 Moderator Analysis: Recall versus Recognition

We found no significant effect of test type (Q(1) = 3.86, *p* = .055). However, since the moderator test was close to significance, we ran effect size analyses of the recall and recognition studies separately. The recall studies showed a medium effect of sleep, with sleep increasing the number of lure words that were falsely recalled as old words, Hedge's *g* = 0.606 (SE = 0.299), which was significantly different from zero (95% CI [0.020, 1.192], *p* = .043). The recognition studies showed a very small non-significant effect in the opposite direction, with sleep reducing the proportion of old responses to lure words, Hedge's *g* = -0.150 (SE = 0.243), indicating no significant difference from zero (95% CI [-0.626, 0.327], *p* = .538).

3.1.2 Moderator Analysis: Number of words

Studies varied in use of either 10, 12, or 15 words within each list. The moderator test indicated a significant effect of number of words (Q(2) = 18.368, *p* < .001). Studies that used 10 words in each list showed a significant increase in the proportion of lure words falsely recalled or recognised as old after sleep than wake (Hedge's *g* = 0.920, SE = 0.193, 95% CI [0.541, 1.300], *p* < .001). No significant effect for 12 words was found (Hedge's *g* = 0.315, SE = 0.302, 95% CI [-0.277, 0.908], *p* = .297). The effect for 15 words was however found to be significantly different from zero (Hedge's *g* = -0.495, SE = 0.165, 95% CI [-0.818, -0.172], *p* = .003), with an *increase* in false memories in the wake group compared to the sleep group.

Since the moderator results also indicated a marginally significant difference in performance between recall and recognition studies, we analysed whether the significant effect of number of words in each list was evident in only those studies using recognition testing. The same effect was confirmed (Q(2) = 22.043, $p < .001$), with lists of 10 words leading to increased false recognition after sleep than wake (Hedge's $g = 0.853$, SE = 0.223, 95% CI [0.417, 1.290], $p < .001$). Lists with 12 words had no effect (Hedge's $g = -0.250$, SE = 0.265, 95% CI [-0.769, 0.269], $p = .345$), whereas word lists with 15 words led to an increase in false recognition in the wake group compared to the sleep group (Hedge's $g = -0.495$, SE = 0.165, 95% CI [-0.818, -0.172], $p = .003$). There was insufficient variation in the list length for the recall test studies to be able to analyse these separately.

3.1.3 Publication Bias

Funnel plots show the distribution of effect sizes around the mean based on sample sizes, with confidence intervals indicating where studies are likely to be positioned, if there is no publication bias. If many studies fall outside the confidence intervals then this indicates that there may be a publication bias (i.e., only studies with larger effect sizes published). Figure 2 shows a funnel plot of effect sizes for proportion of lure words given as old in the sleep versus wake group. An Egger's regression test for funnel plot asymmetry, used for smaller meta-analyses (<25 studies), was run to test for possible publication bias (Egger, Smith, Schneider, & Minder, 1997). There are a number of effect sizes that are outside of the expected distribution; however Egger's regression test indicated no significant funnel plot asymmetry (z = 0.910, $p = .365$) so no evidence for publication bias.

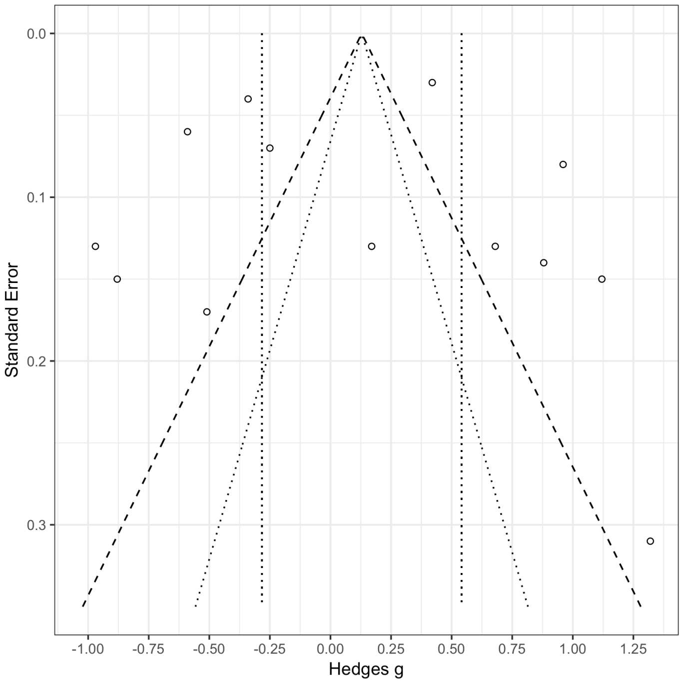


Figure 2.2. Funnel plot showing standard error of the effect size between sleep and wake groups for lure words, with 95% (dotted lines) and 99% (dashed lines) confidence intervals.

3.2 Old Words

The overall Hedge's *g* effect size for old words was 0.159 (SE = 0.126), which again indicated no significant difference from zero (95% CI [-0.088, 0.406], *p* = .206), see Figure 3 for a forest plot of effect sizes. Again, we ran the analysis without the unpublished data, and found no significant change in the effect size, (Hedge's *g* = 0.203, SE = 0.130, 95% CI [-0.052, 0.458], *p* = .458). We therefore continued with the full dataset. There was significant heterogeneity, indicating variance in the data that cannot be explained by the random measurement error, (Q(12) = 28.159, *p* = .005). We therefore again analysed the effects of each of the moderators (see Table 3 for significance of each moderator).

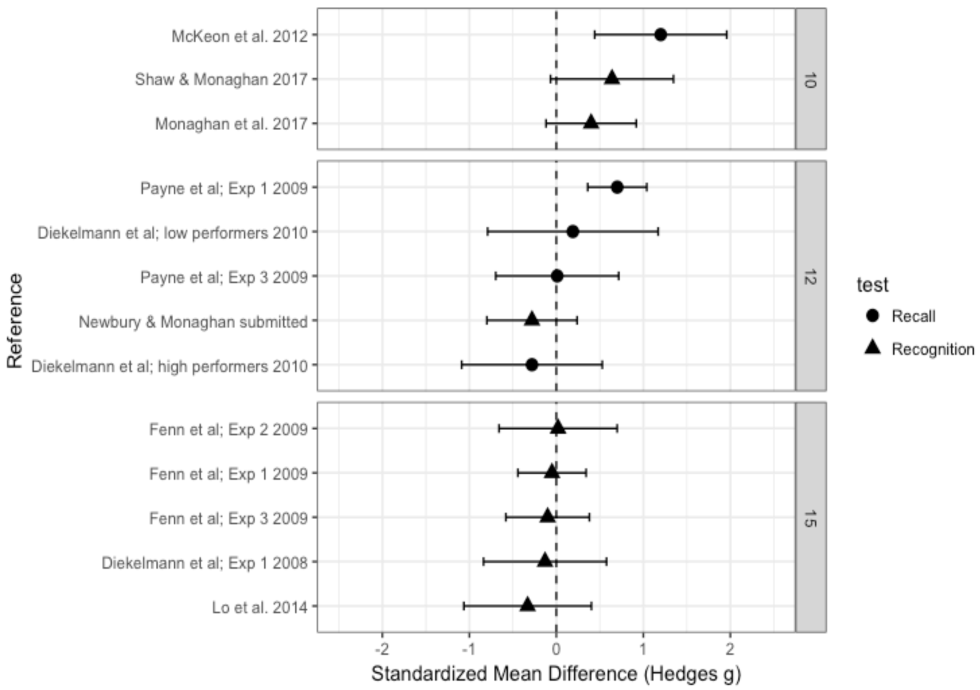


Figure 2.3. A forest plot containing effect sizes and 95% confidence intervals for the difference in the proportion of old words recalled or recognised between sleep and wake groups for old words. Effect sizes further to the right indicate an increase in the proportion of old words accurately recalled or recognised in the sleep group than the wake group.

Table 2.3

The effect of each moderator on the overall effect size difference between sleep and wake groups for old words.

Moderator	*df*	Heterogeneity (Q)	*p*
Recall versus Recognition	1	3.933	.047*
Number of lists	1	1.376	.241
Number of words in each list	2	7.151	.028*
Auditory versus Visual	1	0.088	.767
Emotional versus Neutral	1	0.321	.571
Nap versus Overnight sleep	1	0.259	.611

Note: *p < .05

3.2.1 Moderator Analysis: Recall versus Recognition

Recall versus recognition as a moderator was a significant effect (Q(1) = 3.933, *p* = .047). We therefore ran effect size analyses of recall and recognition studies separately. For studies using a test of recall, we found no significant effect of sleep versus wake (Hedge's *g* = 0.407, SE = 0.256, 95% CI [-0.094, 0.909], *p* = .112), nor was there a significant effect for recognition studies (Hedge's *g* = 0.005, SE = 0.100, 95% CI [-0.190, 0.200], *p* = .958). Therefore, although recall studies significantly differed from recognition studies, with recall studies showing increased performance accuracy after sleep than recognition studies, there was no significant difference in performance accuracy between sleep and wake groups for tests of recall or recognition analysed separately.

3.2.2 Moderator Analysis: Number of words

We found the number of words in each list (10, 12, 15) to be a significant moderating variable (Q(2) = 7.151, *p* = .028). We found a medium effect based on 10 words in the lists (Hedge's *g* = 0.683, SE = 0.230, 95% CI [0.231, 1.134], *p* = .003), with an increase in performance accuracy after sleep compared to wake. We found no significant effect based on 12 words (Hedge's *g* = 0.116, SE = 0.505, 95% CI [-0.334, 0.565], *p* = .614), or 15 words (Hedge's *g* = -0.094, SE = 0.124, 95% CI [-0.338, 0.149], *p* = .448).

Again we tested the effect of number of words for recognition studies only. The same significant effect was found (Q(2) = 6.841, *p* = .033), with lists of 10 words leading to a significant increase in performance accuracy after sleep than wake (Hedge's *g* = 0.484, SE = 0.213, 95% CI [0.066, 0.902], *p* = .023), and no effect based on 12 words (Hedge's *g* = -0.280, SE = 0.265, 95% CI [-0.797, 0.239], *p* = .290), or 15 words (Hedge's *g* = -0.094, SE = 0.124, 95% CI [-0.338, 0.149], *p* = .448).

3.2.3 Publication Bias

Figure 4 shows a funnel plot of effect sizes for accurate recall or recognition of previously seen (old) words in the sleep versus wake group. There are a number of effect sizes that are outside of the expected distribution, however an Egger's regression test indicated no significant funnel plot asymmetry (z = -0.272, *p* = .786), so no evidence of publication bias.

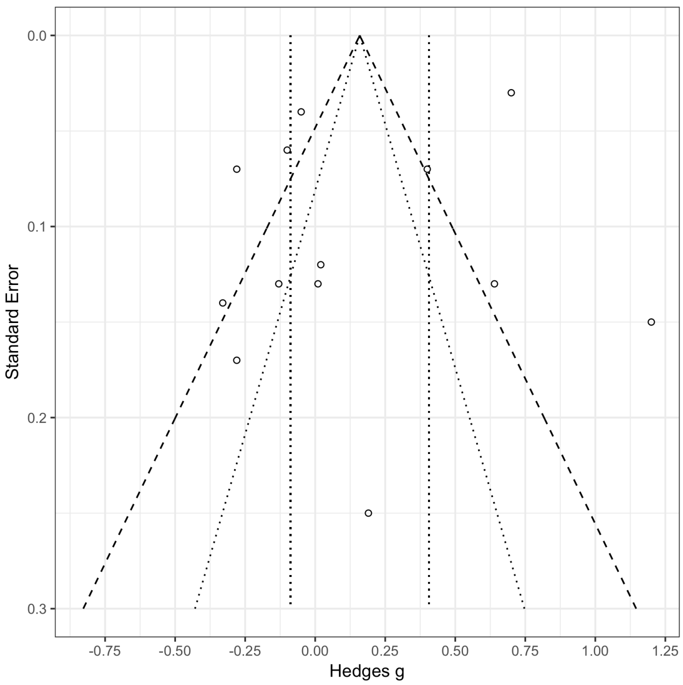


Figure 2.4. Funnel plot showing standard error of the effect size between sleep and wake groups for old words, with 95% (dotted lines) and 99% (dashed lines) confidence intervals.

3.3 New Words

For new words, the overall Hedge's *g* effect size was -0.277 (SE = 0.079), which significantly differs from zero (95% CI [-0.333, -0.022], *p* = .026), suggesting that new words were falsely recalled or recognised as old significantly more in the wake group than the sleep group, see Figure 5 for a forest plot of effect sizes per experiment. Removing the unpublished data did not significantly change the results, (Hedge's *g* = 0.204, SE = 0.083, 95% CI [-0.367, -0.041], *p* = .014) and so we continued with the full dataset. Unlike for lure and old words, heterogeneity was not significant, suggesting that the moderators were not influencing the effect, and any variance in the data can be explained by random measurement error (Q(12) = 7.440, *p* = .827).

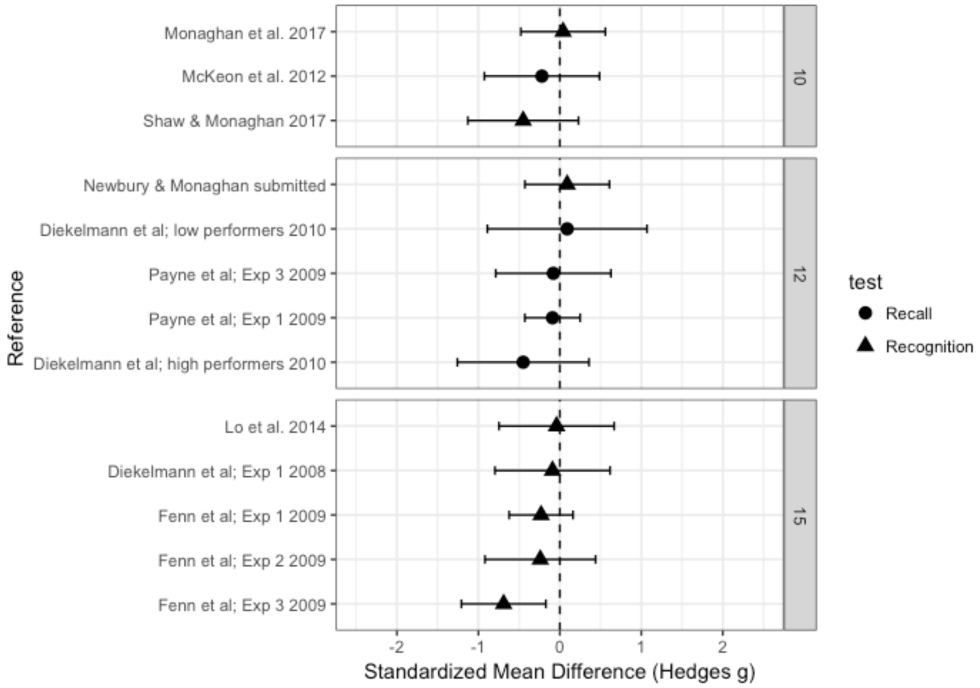


Figure 2.5. A forest plot containing effect sizes and 95% confidence intervals for the difference in the proportion of new words recalled or recognised as old between sleep and wake groups. Effect sizes further to the right indicate an increase in the proportion of new words falsely recalled or recognised as old in the sleep group than the wake group.

3.3.1 Publication Bias

Figure 6 shows a funnel plot of effect sizes for accurate rejection of new words not previously seen for the sleep versus wake groups. There are two effect sizes that are outside of the expected distribution, and an Egger's regression test indicated no significant funnel plot asymmetry (z = -0.179, *p* = .858), so no evidence of publication bias.

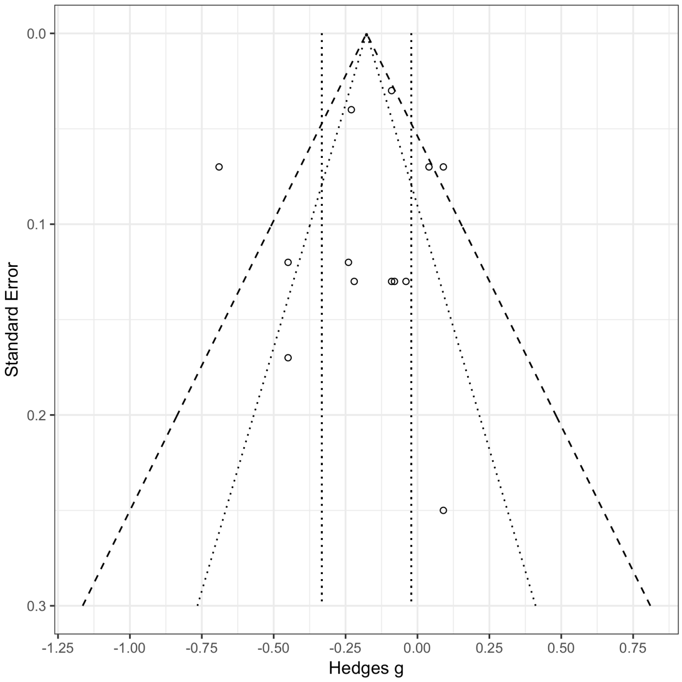


Figure 2.6. Funnel plot showing standard error of the effect size between sleep and wake groups for new words, with 95% (dotted lines) and 99% (dashed lines) confidence intervals.

3.4 Signal Detection Analyses

For those studies in which participants were given a recognition task, we calculated the mean difference between sleep and wake groups in their overall discriminability (d') and response bias (C) for old versus new words (true recognition) and lure versus new words (false recognition). See Table 4 for d’ and C scores per experiment.

Table 2.4

Descriptions of Discriminability (d’) and Response Bias (C) for false recognition (lure versus new words) and true recognition (old versus new words) for sleep and wake groups.

	Discriminability (d')				Response Bias (C)			
	Sleep Groups		Wake Groups		Sleep Groups		Wake Groups	
Authors	False Recog.	True Recog.	False Recog.	True Recog.	False Recog.	True Recog.	False Recog.	True Recog.
Fenn et al. (Experiment 1)	0.819	0.267	0.794	0.216	2.125	2.401	2.067	2.356
Fenn et al. (Experiment 2)	0.986	0.455	0.980	0.362	1.940	2.205	1.856	2.164
Fenn et al. (Experiment 3)	0.927	0.382	0.860	0.274	1.994	2.266	1.884	2.177
Diekelmann et al. (Experiment 1)	0.713	0.188	0.688	0.192	2.076	2.338	2.075	2.323
Lo et al.	0.498	-0.014	0.584	0.000	1.703	1.959	1.651	1.943
Monaghan et al.	0.536	0.174	0.433	0.143	1.819	2.000	1.882	2.027
Newbury & Monaghan	0.389	0.038	0.444	0.094	1.956	2.131	1.954	2.129
Shaw & Monaghan	0.636	0.251	0.374	0.020	1.859	2.051	1.816	1.993

3.4.1 Discriminability (d’)

False Recognition. D-prime (d’) for false recognition was analysed by calculating the z-inverse hit rate (lure words falsely accepted as old/total number of lure words), minus the z-inverse false alarm rate (new words falsely accepted as old/total number of new words) for each experiment.

The overall Hedge’s *g* effect size did not significantly differ from zero (Hedge’s *g* = 0.039, SE = 0.098, 95% CI [-0.153, 0.230], *p* = .692) (see Figure 7 for effect sizes). The test of heterogeneity was not significant, suggesting that there were no potential moderators influencing the result, (Q(7) = 0.736, *p* = .998).

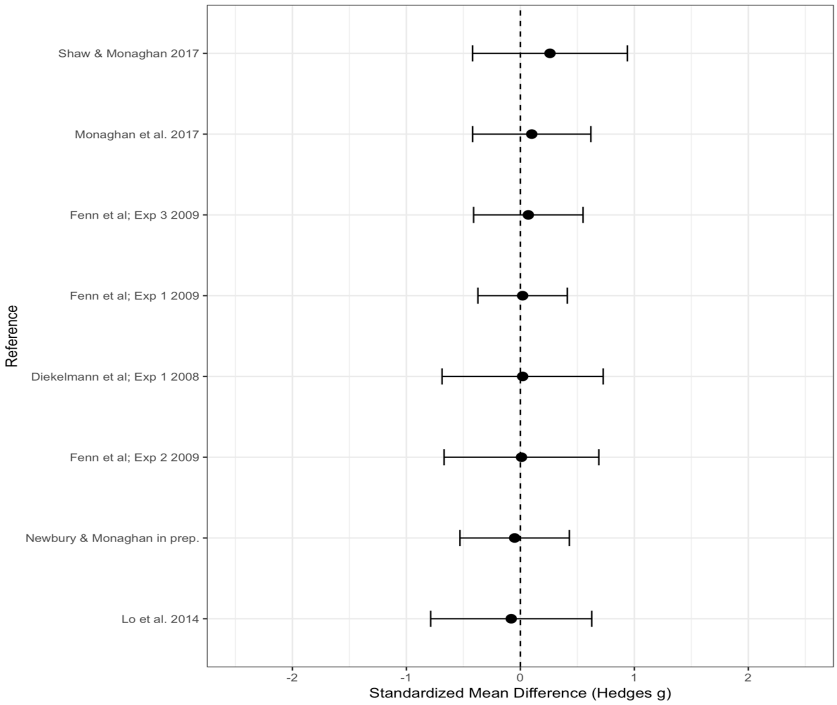


Figure 2.7. A forest plot containing effect sizes and 95% confidence intervals for false recognition discriminability (d') scores. Effect sizes further to the right indicate an increase in discriminability for the sleep group than wake group.

True Recognition. D-prime (d') for true recognition was analysed by calculating the z-inverse hit rate (number of hits/total number of old words), minus the z-inverse false alarm rate (new words falsely accepted as old/total number of new words) for each experiment.

The overall Hedge's $g = -0.044$ (SE = 0.098), which did not significantly differ from zero (95% CI [-0.236, 0.147], $p = .650$) (see Figure 8). The test of heterogeneity was not significant, suggesting that there were no potential moderators influencing the result, ($Q(7) = 4.082$, $p = .770$).

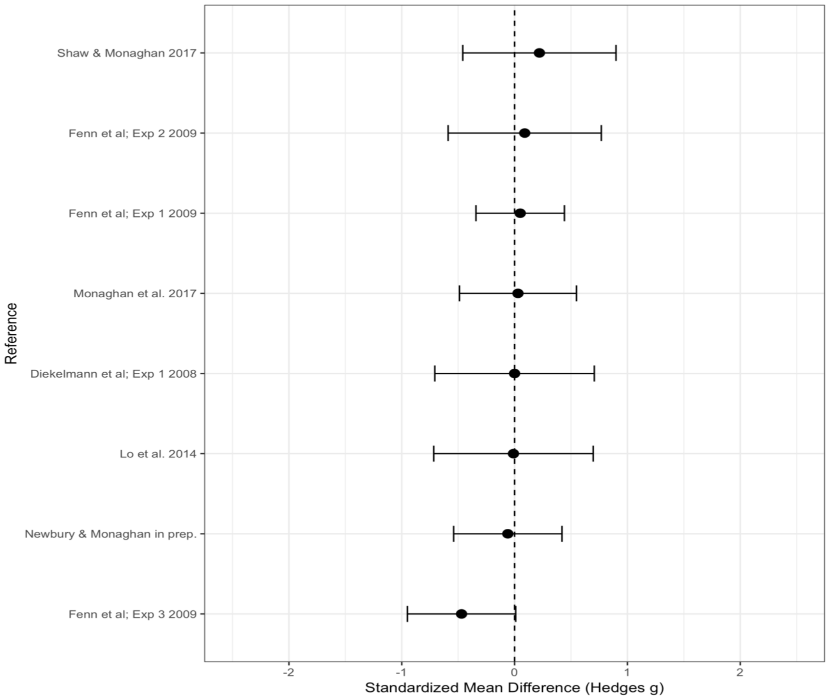


Figure 2.8. A forest plot containing effect sizes and 95% confidence intervals for true recognition discriminability (d') scores. Effect sizes further to the right indicate an increase in discriminability for the sleep group than wake group.

3.4.2 Response Bias (C)

False Recognition. Response bias (C) for false recognition was calculated by the z-inverse transformation of [hit rate (lure words) + false alarm rate]/2. We found no significant effect of sleep on response bias (Hedge's *g* = 0.037, SE = 0.098, 95% CI [-0.155, 0.229], *p* = .706), see Figure 9 for effect sizes. There was no significant heterogeneity, indicating that there were no potential moderators influencing the effect (Q(7) = 0.287, *p* = 1.000).

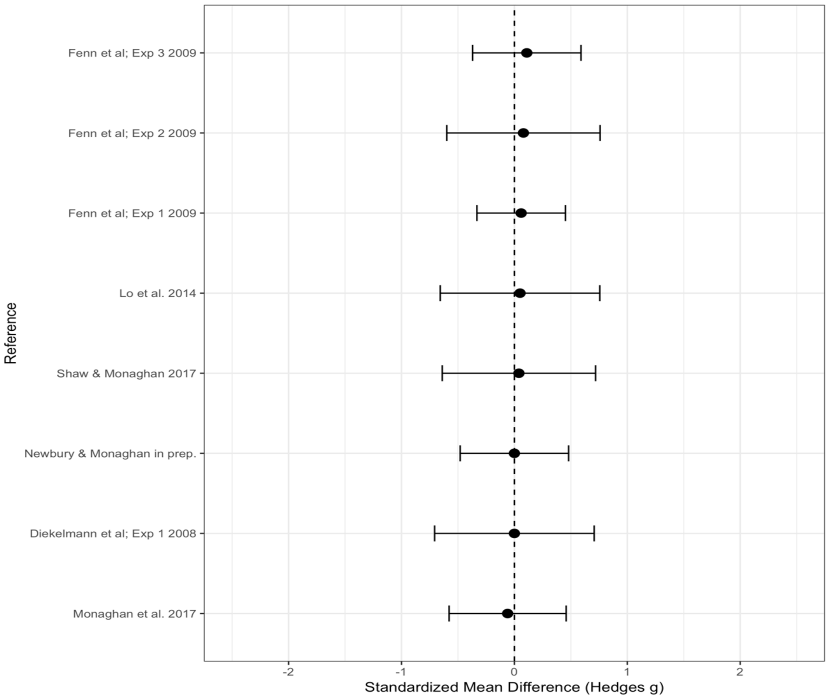


Figure 2.9. A forest plot containing effect sizes and 95% confidence intervals for false recognition response bias (C) scores. Effect sizes further to the right indicate more conservative responses in the sleep group than the wake group.

True Recognition. Response bias (C) for true recognition was calculated by the z-inverse transformation of [hit rate (old words) + false alarm rate]/2. We found no significant effect of sleep on response bias for true recognition, (Hedge's *g* = 0.032, SE = 0.098, 95% CI [-0.159, 0.224], *p* = .741), see Figure 10. There was no significant heterogeneity, indicating that there were no potential moderators influencing the effect (Q(7) = 0.148, *p* = 1.000).

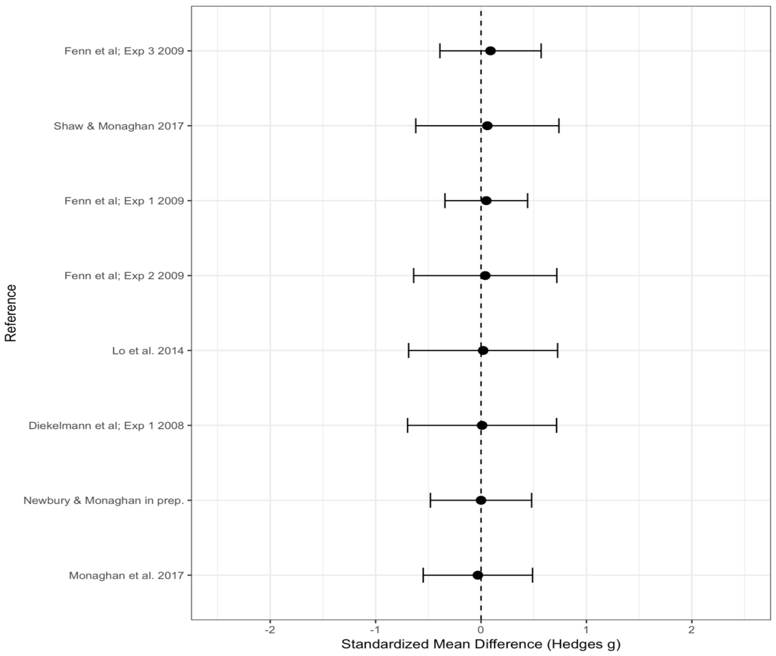


Figure 2.10. A forest plot containing effect sizes and 95% confidence intervals for true recognition response bias (C) scores. Effect sizes further to the right indicate more conservative responses in the sleep group than the wake group.

4. Discussion

The present study examined the effect of sleep on consolidation of seen words, as well as susceptibility to false memories using the DRM procedure. Although we found no overall significant effect of sleep on false memories, the current meta-analysis helps to clarify mixed findings within the literature, by demonstrating that recall versus recognition testing, and shorter list lengths, enhance sleep-based increases in DRM false memories.

Based on the conclusions of a previous meta-analysis conducted by Chatburn et al. (2014), we hypothesised that this lack of an overall effect may have been due to differences between tests of recall and recognition. Based on the previous meta-analysis, as well as the studies presented in the current analysis, we predicted a strong enhancement effect of sleep compared to wake on false recall (Payne et al., 2009). In contrast, for tests of false recognition, sleep has been found to reduce, have no effect, or enhance false memories (Diekelmann et al., 2008; Fenn et al., 2009; Monaghan et al., 2017). A moderator test examining the effect of sleep on false recall and recognition separately found a significant effect of recall, with greater false memories after sleep than wake, whereas recognition tests did not have this same effect. The lack of an effect of sleep on false recognition was further supported by the signal detection analysis, which revealed no significant difference in discriminability or response bias between sleep and wake groups. Thus, this meta-analysis supports Chatburn et al.'s (2014) smaller meta-analysis indicating a significant effect of sleep on false recall, but no effect for false recognition.

However, the larger set of studies investigated in the current meta-analysis enabled us to go further to determine the role of additional task constraints on the effect of sleep on memory. In particular, the results also indicated that list length moderated the effect of sleep on false memories. The studies examined in this analysis used lists consisting of 10, 12 or 15 words. Based on previous research indicating an increase in false memories when more list items were presented, due to increasing associations (Robinson & Roediger, 1997), we predicted two possible hypotheses. If sleep aids in spreading activation of memories equally regardless of the density of the word lists, then we would expect to see no effect of list length on the overall effect size. However, word lists of shorter list length create fewer semantic associations at encoding, thus priming fewer similar, unseen words. If sleep aids memory by increasing the spreading activation in long-term semantic associative memory, then this is more likely to result in activation of lure words for shorter lists where the activation within a network containing fewer semantically related items is sparse compared to a more densely activated network resulting from a longer list (Cai et al., 2009; Sio et al., 2013). The results of the analysis supported this, with an increase in false memories after sleep when studies used lists of 10 words, whereas studies containing lists of 15 words led to a reduction in the proportion of old responses to lure words in the sleep group than the wake group.

For old words, we also found no overall significant difference between the sleep and wake groups on memory performance. This contrasts with previous literature examining the positive effect of sleep on veridical memory consolidation (Dumay & Gaskell, 2007; Plihal & Born, 1997; Wilson et al, 2012). Despite this, two moderators were found to influence the effect size. Tests of recall led to increased performance accuracy after sleep than did tests of recognition for veridical memory of old words. We also found an effect of list length; shorter word lists of ten words led to an increase in accurate memory performance after sleep compared to wake. Therefore, sleep appears to be more beneficial when participants were required to encode fewer words per list. Importantly, this enhancement of sleep effects from short lists for both false and veridical memory was not due to an increase in response bias associated with sleep, as confirmed by the signal detection analyses. The effects were rather specific: only for sparse sets of thematically-related words did sleep improve recognition of old words, and increase acceptance of related but unseen lure words.

For unseen, unrelated (new) words, we expected to see either no difference in performance between sleep and wake groups, due to higher performance accuracy evident in both groups (McKeon, Pace-Schott, & Spencer, 2012; Monaghan et al., 2017; Newbury & Monaghan., submitted), or an increase in accurate rejection of new words after sleep compared to wake due to an overall increase in performance accuracy after sleep (Rasch & Born, 2007). The meta-analysis revealed a small increase in the proportion of new words falsely recalled or recognised as old in the wake group compared to the sleep group. Therefore, the sleep groups were significantly more accurate at rejecting new words as previously seen, supporting previous research indicating a benefit of sleep on accurate memory performance (Davis et al., 2009; Dumay & Gaskell, 2007; Plihal & Born, 1997; Wilson et al., 2012).

Although the current results cannot be extended to apply to general verbal memory consolidation, as the DRM paradigm is designed primarily to examine susceptibility to DRM false memories, and not to investigate veridical memory performance, it should be noted that veridical and false memory within DRM tests are often correlated (e.g., Payne et al., 2009; Shaw & Monaghan, 2017). For those studies that used recall testing and shorter word lists, we saw both an increase in veridical memory, and greater susceptibility to false memories after sleep than after being awake. This, along with the finding that unseen, unrelated new words were rejected more easily by the sleep group, provides support for spreading activation theories of sleep and memory. The current results indicate a role of sleep in associative activation theory (Howe et al., 2009; Roediger et al., 2001), suggesting that shorter word lists with fewer semantic associations benefit from sleep-dependent spreading activation, leading to false acceptance of critical lures to a greater extent than wake, as well as accurate rejection of words with no sematic association.

5. Conclusions

The current meta-analysis of the effects of sleep on veridical and false memory consolidation in DRM tests indicated no overall significant effects. Despite this, it is clear that there are several moderating variables that influence offline memory consolidation. Furthermore, the studies presented in this meta-analysis contain further differences in methodology that may explain the lack of an effect of sleep on both veridical and false memories. For instance, Newbury and Monaghan (submitted) found that sleep improved consolidation of old words to a greater extent than during wake, but only for word lists of negative valence. Monaghan et al., (2017) and Shaw and Monaghan (2017) found evidence for sleep aiding veridical consolidation specifically for those word lists presented to the left hemisphere. Furthermore, Lo, Sim, and Chee (2014) found a reduction in false recognition specifically in older adults, who have previously been found to show different levels of susceptibility to false memories than young adults (Dennis, Kim & Cabeza, 2007; Kensinger & Corkin, 2004) whereas Diekelmann et al. (2010) found an increase in false recall after sleep only for those participants who had an overall low general memory performance. The DRM paradigm does however provide us with evidence for only one type of false memory illusion. Thus, we cannot make firm conclusions regarding the effects of sleep on other forms of veridical and false memories, for example during eye-witness testimony or autobiographical memory for past events.

In conclusion, sleep may therefore improve performance accuracy differentially depending on a number of factors, but the current results do indicate that observations of sleep-enhancement of veridical and false memory effects are task-dependent – potentially sensitive to source monitoring constraints in memory tasks – as well as subject to constraints emergent from the structure of semantic associative memory, as measured by list length, which reflects the density of interconnections within networks of associated words (Monaghan et al., 2017; Robinson & Roediger, 1997). Further investigation as to the effects that these different factors may have on the integration and consolidation of specific information from the short-term to the long-term memory stores will allow for a greater understanding as to the complexities of memory consolidation under different conditions.

References

Adelman, J. S., & Estes, Z. (2013). Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition, 129(3),* 530-535. doi:10.1016/j.cognition.2013.08.014

Bauer, L. M., Olheiser, E. L., Altarriba, J., & Landi, N. (2009). Word type effects in false recall: Concrete, abstract, and emotion word critical lures. *The American Journal of Psychology, 122*(4), 469-481.

Carr, M., & Nielsen, T. (2015). Morning REM sleep naps facilitate broad access to emotional semantic networks. *Sleep*, *38*(3), 433-443. doi:10.5665/sleep.4504

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*(2), 97-111. doi:10.1002/jrsm.12

Cai, D., Mednick, S. A., Harrison, E, M., Kanady, J. C., & Mednick, S. C. (2009). REM, not incubation, improves creativity by priming associative networks. *Proceedings of the National Academy of Sciences, 106*, 10130–10134. doi:10.1073/pnas.0900271106

Chatburn, A., Lushington, K., & Kohler, M. J. (2014). Complex associative memory processing and sleep: A systematic review and meta-analysis of behavioural evidence and underlying EEG mechanisms. *Neuroscience and Biobehavioral Reviews, 47*, 646-655. doi:10.1016/j.neubiorev.2014.10.018

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407-428. doi:10.1037/0033-295X.82.6.407

Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience, 21*(4), 803-820. doi:10.1162/jocn.2009.21059

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience, 11*(2), 114-126. doi:10.1038/nrn2762

Diekelmann, S., Born, J., & Wagner, U. (2010). Sleep enhances false memories depending on general memory performance. *Behavioural Brain Research, 208*(2), 425-429. doi:10.1016/j.bbr.2009.12.021

Diekelmann, S., Landolt, H., Lahl, O., Born, J., & Wagner, U. (2008). Sleep loss produces false memories. *PloS one, 3*(10), e3512-e3512. doi:10.1371/journal.pone.0003512

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18*(1), 35-39. doi:10.1111/j.1467-9280.2007.01845.x

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629-634.

Fenn, K. M., Gallo, D. A., Margoliash, D., Roediger, H. L., & Nusbaum, H. C. (2009). Reduced false memory after sleep. *Learning & Memory, 16*(9), 509-513. doi:10.1101/lm.1500808

Gais, S., & Born, J. (2004). Low Acetylcholine during Slow-Wave Sleep Is Critical for Declarative Memory Consolidation. *Proceedings of the National Academy* *of Sciences of the United States of America,101*(7), 2140-2144. doi:10.1073/pnas.0305404101

Gallo, D. A., & Roediger, H. L. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language, 47*(3), 469-497. doi:10.1016/S0749-596X(02)00013-X

Goldstein, A. N., & Walker, M. P. (2014). The role of sleep in emotional brain function. *Annual review of Clinical Psychology*, *10*, 679-708. doi:10.1146/annurev-clinpsy-032813-153716

Howe, M. L., Candel, I., Otgaar, H., Malone, C., & Wimmer, M. C. (2010). Valence and the development of immediate and long-term false memory illusions. *Memory, 18*(1), 58-75. doi:10.1080/09658210903476514

Howe, M. L., & Wilkinson, S. (2011). Using story contexts to bias children's true and false memories. *Journal of Experimental Child Psychology, 108*(1), 77-95. doi:10.1016/j.jecp.2010.06.009

Howe, M. L., Wimmer, M. C., Gagnon, N., & Plumpton, S. (2009). An associative-activation theory of children's and adults' memory illusions. *Journal of Memory and Language, 60*(2), 229-251. doi:10.1016/j.jml.2008.10.002

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

Kellogg, R. T. (2001). Presentation modality and mode of recall in verbal false memory. *Journal of Experimental Psychology, 27*(4), 913-919. doi:10.1037/0278-7393.27.4.913

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition, 31*(8), 1169-1180. doi:10.3758/BF03195800

Kensinger, E. A., & Corkin, S. (2004). The effects of emotional content and aging on false memories. *Cognitive, Affective, & Behavioral Neuroscience, 4*(1), 1-9. doi:10.3758/CABN.4.1.1

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

Lo, J. C., Sim, S. K. Y., & Chee, M. W. L. (2014). Sleep reduces false memory in healthy older adults. *Sleep, 37*(4), 665-671. doi:10.5665/sleep.3564

Marshall, L., & Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends in Cognitive Sciences, 11*(10), 442-450. doi:10.1016/j.tics.2007.09.001

McDermott, K. B. (1996). The Persistence of False Memories in List Recall. *Journal of Memory and Language, 35*(2), 212-230. doi:10.1006/jmla.1996.0012

McKeon, S., Pace-Schott, E. F., & Spencer, R. M. (2012). Interaction of sleep and emotional content on the production of false memories. *PloS one, 7*(11), e49353. doi:10.1371/journal.pone.0049353

Monaghan, P., Shaw, J. J., Ashworth-Lord, A., & Newbury, C. R. (2017). Hemispheric processing of memory is affected by sleep. *Brain and Language, 167*, 36-43. doi:10.1016/j.bandl.2016.05.003

Newbury, C. R., & Monaghan, P. (2018). *Negative but not Positive Emotional Memories are Enhanced by Sleep.* Manuscript submitted for publication.

Payne, J. D., Schacter, D. L., Propper, R. E., Huang, L., Wamsley, E. J., Tucker, M. A., . . . Stickgold, R. (2009). The role of sleep in false memory formation. *Neurobiology of Learning and Memory, 92*(3), 327-334. doi:doi:10.1016/j.nlm.2009.03.007

Plihal, W., & Born, J. (1997). Effects of Early and Late Nocturnal Sleep on Declarative and Procedural Memory. *Journal of Cognitive Neuroscience, 9*(2), 534-547. doi:10.1162/jocn.1997.9.4.534

Rasch, B., & Born, J. (2007). Maintaining memories by reactivation. *Current Opinion in Neurobiology, 17*(6), 698-703. doi:10.1016/j.conb.2007.11.007

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

Robinson, K. J., & Roediger, H. L. (1997). Associative Processes in False Recall and False Recognition. *Psychological science, 8*(3), 231-237. doi:10.1111/j.1467-9280.1997.tb00417.x

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology, 21*(4), 803-814. doi:10.1037/0278-7393.21.4.803

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8*(3), 385-407. doi:10.3758/BF03196177

Sharkawy, J. E., Groth, K., Vetter, C., Beraldi, A., & Fast, K. (2008). False memories of emotional and neutral words. *Behavioural Neurology, 19*(1), 7-11.

Shaw, J. J., & Monaghan, P. (2017). Lateralised sleep spindles relate to false memory generation. *Neuropsychologia*, *107*, 60-67. doi:10.1016/j.neuropsychologia.2017.11.002

Sio, U., Monaghan, P., & Ormerod, T. (2013). Sleep on it, but only if it is difficult: Effects of sleep on problem solving. *Memory & Cognition, 41*(2), 159-166. doi:10.3758/s13421-012-0256-7

Smith, R. E., & Engle, R. W. (2011). Study Modality and False Recall. *Experimental Psychology, 58*(2), 117-124. doi:10.1027/1618-3169/a000076

Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin & Review, 5*(4), 710-715. doi:10.3758/BF03208850

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature Cell Biology, 437*(7063), 1272-1278. doi:10.1038/nature04286

Stickgold, R., & Walker, M. (2004). To sleep, perchance to gain creative insight? *Trends in Cognitive Sciences, 8*(5), 191-192. doi:10.1016/j.tics.2004.03.003

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software, 36*(3). doi:10.18637/jss.v036.i03

Walker, M. P., & Stickgold, R. (2006). Sleep, memory, and plasticity. *Annual Review of Psychology, 57*, 139-166. doi:10.1146/annurev.psych.56.091103.070307

Watson, J. M., McDermott, K. B., & Balota, D. A. (2004). Attempting to avoid false memories in the Deese/Roediger—McDermott paradigm: Assessing the combined influence of practice and warnings in young and old adults. *Memory & Cognition, 32*(1), 135-141. doi:10.3758/BF03195826

Wilson, J., Baran, B., Pace-Schott, E. F., Ivry, R. B., & Spencer, R. M. C. (2012). Sleep modulates word-pair learning but not motor sequence learning in healthy older adults. *Neurobiology of Aging, 33*(5), 991-1000. doi:10.1016/j.neurobiolaging.2011.06.029

Chapter 3: Lateralised Encoding of DRM Word Lists Does Not Affect Memory Processing.

The previous chapter provided a meta-analytic review of previous literature on sleep and DRM memories, establishing a role of sleep in the formation of both veridical and false memories dependent on moderating factors. We demonstrated an effect of recall versus recognition testing and number of words within each word list on sleep-dependent veridical and false memories.

The following chapter aimed to empirically investigate the effects of lateralisation of memories on veridical and false memories. Previous literature has demonstrated a LH bias for veridical memory, and a RH bias for false memory. Furthermore, findings suggest that sleep boosts this lateralisation effect, however previous investigations have focused on effects of lateralisation at retrieval. We sought to examine whether this sleep effect can be extended to processes of encoding.

Statement of Author Contribution.

In the chapter entitled "Lateralised Encoding of DRM Word Lists Does Not Affect Memory Processing", the authors agree to the following contributions:

Chloe R. Newbury – 80% (Experimental design, data collection, data analysis, and writing)

Signed: C Newbury Date: 25/04/2019

Professor Padraic Monaghan – 20% (Experimental design and review)

Signed: P Mon Date: 25/04/2019

Abstract

Previous research indicates that the left hemisphere (LH) activates closely associated words with fine semantic networks, whereas the right hemisphere (RH)activates broader semantic networks. In studies of lateralisation effects at retrieval, DRM studies support this, indicating increased accurate word recognition in the LH, and increased false recognition of associated words in the RH. Sleep has been found to promote this LH advantage for accurate memories, as well as increase spreading activation of semantic networks across hemispheres, increasing false memories. The current study examines whether these effects are replicated when word lists are presented laterally at encoding rather than retrieval. In Experiment 1, participants viewed lists of semantically related words to either the LH or RH. After a 5-minute delay, participants were presented with either previously seen, unseen related, or unseen unrelated words, and rated how confident they were in recognising the word. In Experiment 2, participants completed the testing session after a 12-hour period of overnight sleep or daytime wake. The results indicate no difference in accurate or false recognition when words were encoded in either hemisphere, and no effect of sleep on memory performance. Thus, hemispheric differences in memory performance are specific to when words are lateralised during recognition.

1. Introduction

Lateralised visual half field tasks, whereby stimuli presented to the right visual field (RVF) results in left hemisphere (LH) involvement, and stimuli presented to the left visual field (LVF) results in right hemisphere (RH) involvement, have indicated that there is a LH advantage for word/language processing (Ellis, 2004; Pirozzolo & Rayner, 1977). Despite this LH advantage, both the LH and RH have been found to play distinctive roles in the processing of language. Semantic priming studies have found that the priming of distantly related words is greater for those words presented to the RH than the LH (Chiarello, Burgess, Richards, & Pollock, 1990), and Beeman, Friedman, Grafman, Perez, Diamond, and Lindsay (1994) found that summation priming of semantically associated words was greater in the RH when target words were distantly related, whereas priming of more closely related words was greater in the LH. These lateralisation studies therefore suggest a LH bias for the activation of close semantic networks that process distinctly related words, whereas the RH activates broader semantic networks with more distantly related meanings of associated words (Beeman & Bowden, 2000; Beeman & Chiarello, 1998; Burgess & Simpson, 1988; Monaghan, Shillcock, & McDonald, 2004). This has been termed the fine versus coarse semantic coding hypothesis.

Lateralisation studies of the Deese-Roediger-McDermott (DRM) paradigm have been used to investigate these asymmetries in the semantic processing of word lists in the LH and RH. In tests using the DRM paradigm, participants are exposed to lists of words that are semantically related (e.g. *dream, bed, night, snooze etc.*). During testing, participants are asked to recall words from each list, or are given a recognition task in which they must decide whether each word they are presented with was present in the initial lists. Words are categorised as either previously seen in the initial word lists (old words), related but unseen words (lure words), or unrelated, unseen words (new words). During tests of recall and recognition, participants are likely to falsely remember being presented with the lure words, indicating a false memory. Based on the fine versus coarse semantic coding hypothesis, whereby there is a broader spread of activation of semantic associates for words processed in the RH than LH, it would be expected that the RH would produce higher rates of false recognition of lure words, and the LH would be more accurate at discriminating old from lure words. A range of studies presenting word lists centrally during word learning, and to the LVF/RH or RVF/LH during subsequent testing, support this view (Bellamy & Shillcock, 2007; Faust, Ben-Artzi, & Harel, 2008; Ito, 2001; Westerberg & Marsolek, 2003), with findings suggesting a LH advantage in correctly rejecting lure words, and a RH bias in the processing of false memories.

A range of evidence suggests that sleep affects both declarative and procedural memory compared to daytime wake (Marshall & Born, 2007; Rasch & Born, 2013; Stickgold & Walker, 2005; Walker & Stickgold, 2006). During sleep, reactivation of memories initially stored short-term in the hippocampus are integrated into the long-term store in the neo-cortex (Marshall & Born, 2007; Rasch & Born, 2013). This process leads to stronger semantic associations, thus improving veridical memory performance, but also leads to increased memory for words with related meanings, termed false memories (Diekelmann, Born & Wagner, 2010; Monaghan, Shaw, Ashworth-Lord & Newbury, 2017; Payne et al., 2009). However, a recent meta-analysis conducted by Newbury and Monaghan (2018), found no overall effect of sleep on either veridical or false memories, although differences between recall versus recognition testing, and the number of words per DRM list did moderate this effect. Despite this, previous work does suggest that sleep leads to the reactivation of new memories, integrating them with previously learned information (Cai, Mednick, Harrison, Kanady, & Mednick, 2009; Sio, Monaghan, & Ormerod, 2013). Indeed, Tamminen, Payne, Stickgold, Wamsley, and Gaskell (2010) found that sleep supported the integration of newly learned words into the mental lexicon, highlighted by lexical competition effects between novel and previously known words. This competition effect was not evident immediately after exposure or after a period of wakefulness, consistent with the hypothesis of an integration of knowledge from hippocampal to neocortical systems.

The effect of sleep on the hemispheric processing of both veridical and false memories is however still under investigation. Evidence suggests that during the first half of a night of sleep, the RH is more active than the LH (Bolduc, Daoust, Limoges, Braun, & Godbout, 2003; Casagrande & Bertini, 2008; Casagrande, Violani, De Gennaro, Braibanti & Bertini, 1995). Thus, an increase in DRM false memories after sleep may be due to RH broader semantic associations, resulting in an increase in false recognition after sleep specifically in the RH, rather than a general increase in false recognition. If sleep does facilitate the reactivation of memories from hippocampal systems to the long-term store in the hippocampus, we would also expect a shift to a LH advantage for accurate memory after sleep. Monaghan et al. (2017) tested these hypotheses, presenting words to the LH or RH during recognition, after a period of overnight sleep or daytime wake. They found that sleep led to an increase in accurate recognition of words presented in the LH, supporting the hypothesis that sleep reactivates new memories in the hippocampus, leading to consolidation of information into existing memories in the long-term store. However, they found no differential effect of sleep on LH versus RH false memories. It is thus suggested that sleep leads to enhanced spreading activation equally in the two hemispheres, rather than a specific

RH bias for false memories after sleep. However, in a daytime nap paradigm, Shaw and Monaghan (2017) did find a RH bias for false memories after sleep, and this effect was correlated with stage 2 sleep spindles specifically in the RH.

Despite this evidence for differences between LH and RH memory processing after sleep, there has been little investigation into the effects of processing words in the RH or LH during initial encoding of the word lists, rather than during retrieval. According to two theories of DRM false memories, namely the fuzzy trace theory and spreading activation theory (Kim & Cabeza, 2007), false memories occur due to memory for the gist, or activation of semantically similar information. False memories occur at the point of recognition due to the activation monitoring framework (Gallo, 2010), such that recalling or recognizing previously seen information activates semantically similar memories or the gist of that memory. However, false memories may also be influenced at the level of encoding, when the semantic knowledge is learned and memory for the gist can be encoded (Kim & Cabeza, 2007; see Straube, 2012 for a review of encoding and retrieval processes). Indeed, specific brain regions that are active during encoding have been linked to the formation of false memories, (Kim & Cabeza, 2006), suggesting an influence of encoding on both consolidation and retrieval of both veridical and false memories.

In one study by Fabiani, Stadler, and Wessels (2000), DRM lists were presented to either the LH or RH during encoding. Although there was no significant difference in performance accuracy, there was a significant difference in response times (RT) to lure words that were presented to the LH versus RH. Lure words whose associated lists were presented in the RH were responded to significantly slower than those in the LH. This finding suggests that word lists encoded in the RH elicited a broader spread of activation, and so participants took longer to access the information needed to make an old/new judgement. If this is the case, this study demonstrates a RH bias in the formation of false memories specifically during encoding of DRM lists. Since the accuracy data did not correspond to the RT data, it may be that the effect of encoding in the RH or LH on the formation of false memories is not as strong as the hemispheric differences present during memory retrieval. Furthermore, this study presented word lists to the LH or RH twice during encoding, which could have influenced how the information is processed, and may have led to the transfer of word lists to both hemispheres. This could provide another explanation as to why inconsistencies between accuracy and RT data were found. Indeed, previous work investigating hippocampal activation during initial encoding of stimuli suggests no hemispheric asymmetries in the encoding of stimuli (Hocking, McMahon & De Zubicaray, 2009). Thus, it is unclear what effect encoding of word lists in the LH versus RH would have on subsequent memory, and whether sleep would differentially effect LH versus RH memory.

We therefore aimed to further investigate whether any differences in LH and RH encoding leads to differences in performance, and whether sleep plays a role in the processing of these memories, using DRM tests of false recognition. We presented DRM word lists one time only during encoding, to either the LH or RH. In Experiment 1, participants were then given a short break before taking part in the recognition task. In Experiment 2, participants came back 12 hours later, after either daytime wake or overnight sleep. If lateralised encoding of DRM lists affects performance, we would expect a LH bias in the encoding of veridical memories, and a RH bias in the encoding of false memories, with this effect enhanced after overnight sleep, following a similar pattern to DRM studies investigating hemispheric differences at retrieval (Monaghan et al., 2017; Shaw & Monaghan, 2017). However, the results may also demonstrate no difference between the LH and RH in encoding of word lists, suggesting that differential effects of sleep on RH and LH processing are specific to retrieval processes.

2.1 Experiment 1: Effects of lateralisation at encoding on the processing of DRM word lists

In this study, we presented participants with DRM word lists to either the RVF/LH or LVF/RH during initial presentation of the word lists. After a 5-minute delay, we then tested participants on their recognition of previously seen (old) words, unseen related (lure) words, and unseen unrelated (new) words. If lateralisation effects are evident during encoding of the word lists, then we would expect a LH advantage for previously seen words, and a RH bias for unseen, related lure words. However, if encoding the word lists in either hemisphere has no effect on retrieval processes, then we would expect no difference in performance between the two hemispheres.

2.1.1 Methods

2.1.1.1 Participants

Forty-eight participants took part in the study for course credit through Lancaster University's research participation system. All participants gave informed consent and were fully debriefed at the end of the study. Participants who were left handed or ambidextrous (*n* = 12), based on the Edinburgh Handedness Questionnaire (Oldfield, 1971) were removed from the analysis, as well as those who were non-native English speakers (*n* = 6). Therefore, 31 participants (25 female) with a mean age of 19.74 years were included in the final analyses.

2.1.1.2 Materials

DRM word lists and critical lures were taken from lists used by Monaghan et al., (2017). The same twelve lists of 10 words were selected, corresponding to the critical lures: *car, chair, doctor, bread, fruit, sleep, thief, river, needle, music, mountain, king.* These DRM word lists were originally taken from Stadler et al's (1999) norms, and lists were selected that corresponded to a lure word that resulted in false memories in the range of 30-70%. The 10 words in each list were then randomly selected from the 15 words in Stadler et al's (1999) original lists.

During testing 48 words were used, two previously seen from each group (total of 24), the lure word for each group (total of 12) and 12 unseen, unrelated words. Unrelated words were also the same as those used in Monaghan et al., (2017). All words were presented on the screen in Courier New, bold, 18-point, black font, and lower case.

2.1.1.3 Procedure

During the training session, participants were seated approximately 65cm from the computer screen and were instructed to remember the words that appeared on the screen. Participants were presented with a central fixation point for 500ms before words from one DRM word list were presented one at a time to the left or right of the screen for 120ms. This production time was chosen as it is long enough for participants to see each word, but not long enough for eye movement to occur, which would cause a crossing of information to the other hemisphere. Words were presented so that the middle letter was five characters from the fixation point, equating 3cm to the left or right of the fixation point. All words from one list were presented to the same side of the fixation point, and the presentation of each list to the left or right of the screen was randomised for each participant. Participants were instructed to focus their gaze on the central fixation point throughout and not look to the left or right of the fixation point. A Tobii eye tracker was used to ensure that participants followed these instructions correctly and remained fixated on the central cross during word presentation.

In between each list of words, participants were asked to complete three maths problems that were presented for 10s each. These maths problems were used as a distractor task, in order to prevent rehearsal of the word lists. Each maths problem was only presented once throughout the training phase and was presented in random order for each participant. After participants had attempted 3 maths problems, the fixation cross reappeared and participants were presented with another DRM word list. The order in which DRM word lists were presented was randomised, as was the order in which each word in each list appeared.

Participants then completed the Edinburgh Handedness Questionnaire before taking part in the second part of the study. During the testing phase, participants were presented with a central fixation cross for 500ms, before a test word was presented to them in the centre of the screen. The presentation of the words was randomised, and each word was present on the screen until participants made their decision. Participants were asked to decide whether they recognised the word from the initial training phase or not. They were asked to rate on a scale from 1 to 6 how confident they were in their judgement of whether each word was an old word (previously seen) or a new word (not previously seen), with 1 indicating definitely old, and 6 indicating definitely new. Once participants made their decision, they were presented with another fixation point for 500ms, followed by another word in the centre of the screen.

2.1.1.4 Data Analysis

We performed a series of Linear Mixed-Effects Models (Baayen, 2008; Jaeger, 2008) in R (R Core Team, 2015) to model both response accuracy and response time (RT). For both accuracy and RT we took a hypothesis-driven approach to model building, such that a model was initially fitted specifying the random effects of participants and items *(words)*. We then considered the effect of word type (old/lure/new) and hemisphere (left/right) on model fit, before adding the interaction effect word type by hemisphere. We ran likelihood ratio test comparisons to compare models, to indicate whether the inclusion of each factor and the interaction term significantly improved model fit. Finally, we assessed whether inclusion of random effects in the model were justified. For accuracy we also conducted a Bayesian analysis, to analyse the likelihood that the results provide strong evidence for the null hypothesis (H0) or the alternative hypothesis (H1).

We also analysed eye tracking data, to assess whether participants were fixated on the centre of the screen as per their instructions, to investigate whether this had an influence on any effects. To do this, each word list that participants viewed was given a score from 0-10, based on how many of the words in that list that participants were correctly fixating to the centre of the screen for. If participants were correctly fixated to the centre for each of the 10 words in a list, then they would receive a score of 10 for that list. Each of the two previously seen words, and the unseen, related lure word associated with each list was then given this same score.

Finally, we also analysed participants' confidence in their responses. First, we analysed overall confidence ratings for responses to old and lure words. Confidence ratings ranged from 1 (confident that the word was old) to 6 (confident that the word was new). We also analysed confidence ratings for only the old and lure words that were responded to as old (and so with a confidence rating of 1-3). As with the accuracy and RT data, we initially fitted a model with random effects of participants and items, and then built up the model by first adding main effects of word type, hemisphere, and eye-tracking score, and then adding the interaction terms.

2.1.2 Results

2.1.2.1 Accuracy

We performed a series of Generalized Linear Mixed-Effects Models, modelling the probability (log odds) of response accuracy. For previously seen old words, the accuracy measure represented the proportion of responses correctly accepted as previously seen. For unseen related and unrelated words (both lure and new words), the accuracy measure represented the proportion of responses correctly rejected as previously seen. Thus, for unseen related lure words, lower accuracy was equal to an increase in false recognition. See Figure 1 for mean recognition accuracy across conditions.

Adding the fixed effect of word type to a model with only random effects significantly improved model fit, ($\chi^2(2) = 51.619$, $p = 6.181 \times 10^{-12}$). As predicted, the results indicated that lure words were responded to significantly less accurately than both old words and new words, indicating that false memory effects were observed for unseen, related lure words. The inclusion of the effect of hemisphere did not significantly improve model fit compared to a model with the same random effects plus the main effect of word type, ($\chi^2(1) = .031$, $p = .859$), indicating no difference in memory performance dependent on whether word lists were encoded in the LH or RH.

Similarly, the interaction term word type by hemisphere did not improve model fit, ($\chi^2(3) = .955$, $p = .812$).

We then analysed whether the inclusion of random effects was justified in the final model. Likelihood ratio tests with the same fixed effect of word type but varying in random effects indicated that the inclusion of random effect of items on intercepts was justified. There was a significant difference between a model including random effects of both participants and items on intercepts and that including only the participant effect, ($\chi^2(1) = 33.217$, $p = 8.243 \times 10^{-9}$). However, the inclusion of the random effect of participants on intercepts was not justified. There was no significant difference between a model including random effects of both participants and items on intercepts, and one including only the item effect, ($\chi^2(1) = 1.346$, $p = .246$).

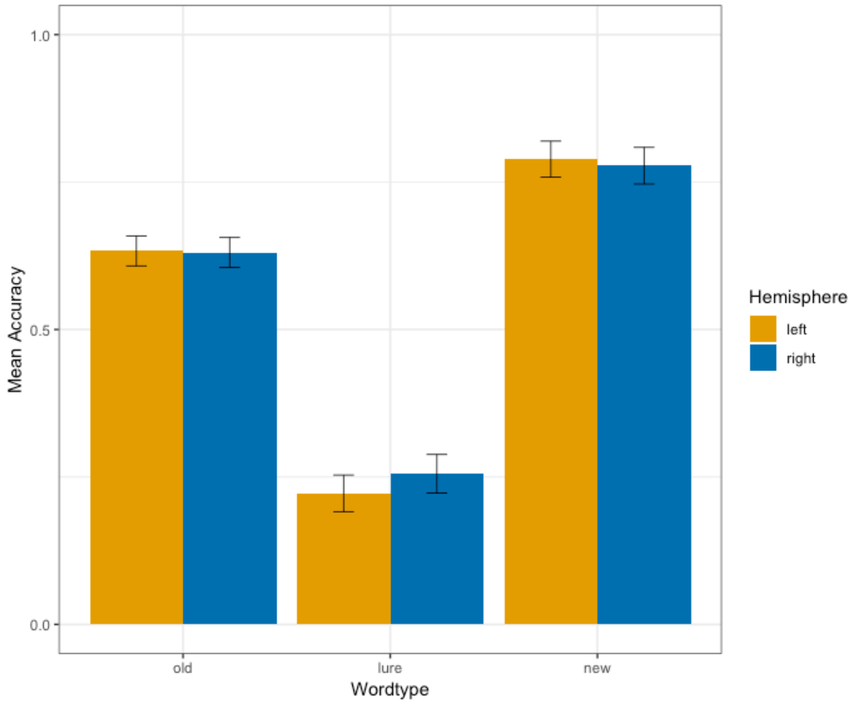


Figure 3.1. Mean accuracy performance for old, lure, and new words in the left and right hemisphere, error bars show ±1 SEM.

Since the results of this study suggest that there is no effect of hemispheric asymmetry on false memories when word lists are lateralised at encoding, rather than at retrieval, we also conducted a Bayesian analysis, to analyse the likelihood that the results provide strong evidence for the null hypothesis (H0) or the alternative hypothesis (H1). Previous research presenting words to the RVF/LH and LVF/RH at the recognition phase have found a RH bias for false memories. We therefore used the results from Bellamy and Shillcock (2007) to set our prior assumptions. They found that unseen, related lure words were significantly more likely to be falsely recognised in the RH than the LH. We chose the results from this paper as they provided adequate data for us to compare to the current study (means and F values). A value greater than 1 suggests support for H1, and thus support previous research, whereas values less than 1 indicate support for H0. Using guidelines from Dienes and McLatchie (2017) for the analysis, our results revealed support for the null hypothesis, $B_{\mathrm{H}(-0.0747,\ 0.037)} = 0.112$. As a convention, Bayes factors < 0.33 are considered strong evidence for the null hypothesis (Jeffreys, 1939/1961). Therefore, these results suggest that when words are lateralised at encoding, rather than at retrieval, there is no evidence for a RH bias for false memories.

We then analysed the eye tracking data, to assess whether this had an influence on the overall lack of an effect. See the Appendix for mean eye-tracking score across conditions. Using Generalized Linear Mixed-Effects Models, we analysed whether, for old and lure words, eye tracking score influenced the results found. Adding the main effect of eye tracking score to a model with only random effects and the significant effect of word type did not improve model fit, ($\chi^2(1) = .409$, $p = .522$), nor did adding a word type by eye tracking interaction ($\chi^2(2) = 1.139$, $p = .569$), hemisphere by eye tracking interaction ($\chi^2(3) = .570$, $p = .903$), or word type by hemisphere by eye tracking interaction ($\chi^2(6) = 7.236$, $p = .230$).

2.1.2.2 Confidence Ratings

For previously seen versus unseen, related lure words, we were also interested in participants confidence in their judgement. See Figure 2 for mean confidence ratings for old and lure words in both the left and right hemisphere. Using Linear Mixed-Effects Modelling, we found that adding the main effect of word type significantly improved model fit compared to a model with only the random effects of participants and items, ($\chi^2(1) = 4.816$, $p = .028$). Lure words were falsely accepted with more confidence than old words were accurately accepted, (estimate = .406, SE = .179, $t = 2.270$). Adding the main effect of hemisphere did not improve model fit ($\chi^2(1) = .879$, $p = .349$), nor did the main effect of eye tracking score ($p = .538$).

We further analysed confidence ratings for only the old and lure words that were responded to as old. We again used Linear Mixed-Effects Modelling, starting with a model including only the random effects of participants and items, and then adding the main effects of word type, hemisphere, and the interaction between the two. We found that adding word type did not significantly improve model fit compared to a model with only random effects of participants and items, ($\chi^2(1) = .391$, $p = .532$), indicating no difference in confidence of accepting previously seen, and related, unseen words as old. We also found no main effect of hemisphere ($\chi^2(1) = .557$, $p = .456$), eye tracking score ($p = .200$) or any of the interaction terms ($p > .05$).

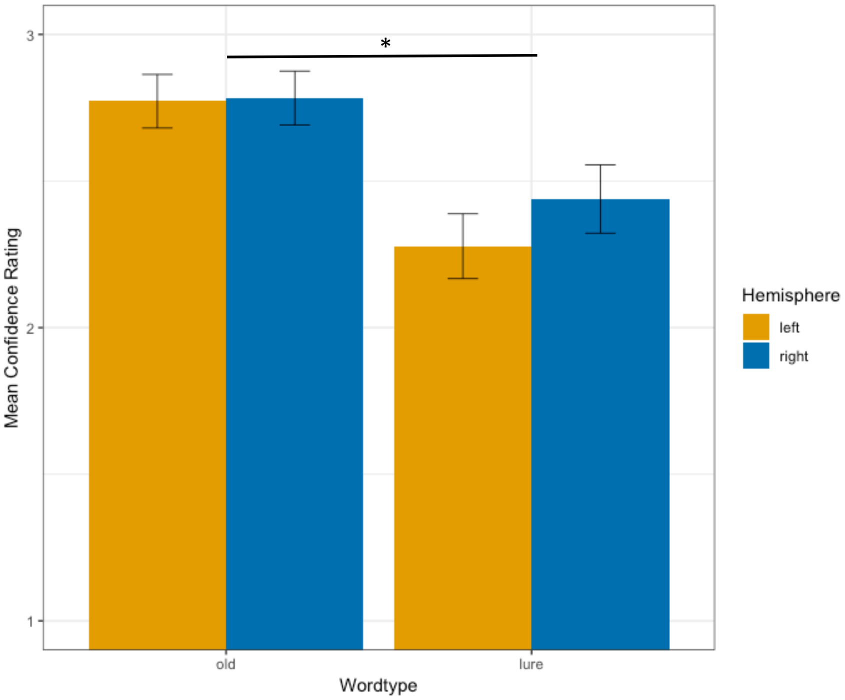


Figure 3.2. Mean confidence ratings for old and lure words in the left and right hemisphere, error bars show ±1 SEM, **p* < .05.

2.1.2.3 Response Time

Response times (RT) for correct responses were analysed to test whether there were any speed-accuracy trade-off effects that may have influenced the accuracy findings, as well as to establish whether there were any asymmetries in participants' speed of access to stored information. Response times exceeding 2.5 standard deviations from the mean were removed from the analysis, leading to the removal of 1.794% of the data, and leaving 836 data points. Based on the suggestions from Baayen, Feldman, and Schreuder (2006), we reduced skew in the RT distribution by transforming latencies to $\log_{10}$(RT). We analysed the logRTs using Linear Mixed-Effects Modelling, considering the random effects of both participants and items on intercepts. We ran a series of models with the same random effects structure and varying in fixed effects structure. We compared models using likelihood ratio tests, using the REML = FALSE setting. See Figure 3 for mean response times across conditions.

The results followed the same pattern as the accuracy data. Adding the main effect of word type to a model with only random effects significantly improved model fit, ($\chi^2(2) = 27.398$, $p = 1.124 \times 10^{-6}$). Adding the main effect of hemisphere to a model with random effects and the main effect of word type did not significantly improve model fit, ($\chi^2(1) = 0.976$, $p = .323$), nor did adding the interaction term word type by hemisphere, ($\chi^2(3) = 1.450$, $p = .694$).

Similar to the accuracy data, we then analysed whether the inclusion of random effects was justified in the final model. There was a significant difference between a model including random effects of both participants and items on intercepts, and one including only a random effect of participants, ($\chi^2(1) = 3.842$, $p = .049$), or only a random effect of items, ($\chi^2(1) = 211.800$, $p = 2.2 \times 10^{-16}$). The inclusion of the random effect of participants and items on intercepts was therefore justified.

We then assessed whether eye tracking score had any influence on the lack of any hemispheric difference for old and lure words. Again, adding eye tracking score to a model with random effects and main effect of word type did not improve model fit, ($\chi^2(1) = 1.625$, $p = .202$), nor did adding any interaction terms ($p > .05$).

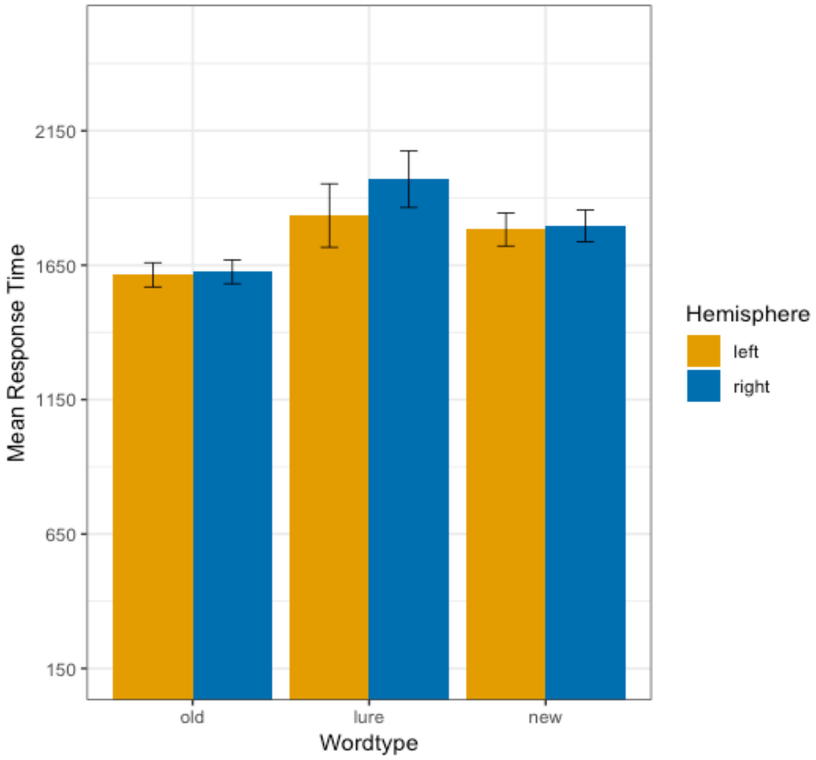


Figure 3.3. Mean response times for old, lure, and new words in the left and right hemisphere, error bars show ±1 SEM.

2.2 Experiment 2: The effect of sleep on the encoding of DRM word lists in the LH or RH.

In Experiment 2, we aimed to establish whether encoding word lists in the LH or RH led to differences in memory consolidation after sleep, thus affecting the way that memories are reactivated and stored long term. If asymmetries in encoding word lists leads to differences in the reactivation of memories, then we would expect to see a beneficial effect of sleep on veridical memory performance for word lists encoded in the LH, and an increase in false acceptance of words related to lists encoded in the RH. However, previous research investigating hippocampal activation during encoding of stimuli found no hemispheric asymmetries (Hocking et al., 2009). If this is the case, then reactivation of new information in hippocampal systems will not be affected by asymmetries in encoding of the word lists, and we would expect no difference in either veridical or false recognition in the LH versus RH after sleep.

2.2.1 Methods

2.2.1.1 Participants

Sixty-eight participants took part in the study for course credit through Lancaster University's research participation system. All participants gave informed consent and were fully debriefed at the end of the study. Exclusion criteria set out before data collection meant that a further 18 participants were tested but removed from the final analysis due to being left handed or ambidextrous (*n* = 7), non-native English, (*n* = 5), not sleeping for 6 hours in between sessions in the sleep group (*n* = 4), napping throughout the day in the wake group (*n* = 3), and one participant did not return for the second session.

2.2.1.2 Materials

The same stimuli as in Experiment 1 were used in this study. Participants were also asked to complete a short questionnaire on their sleep habits to ensure a regular sleep schedule.

2.2.1.3 Procedure

Participants were randomly allocated to either the sleep (*n* = 35) or wake group (*n* = 33) condition. Those in the sleep group were trained on DRM word lists at 9pm and tested at 9am the following day, and those in the wake group were trained on the stimuli at 9am and tested the same day at 9pm. Participants in the sleep group were asked to get a minimum of six hours of sleep in between sessions, and this was assessed by the sleep habits questionnaire given to participants, and confirmed with Actigraph monitors that participants wore overnight. Those in the wake group were required not to nap between sessions, assessed through the sleep habits questionnaire. At the first session, participants first completed the handedness questionnaire and sleep habits questionnaire. They then saw the 12 lists of 10 words presented to either the RVF or LVF, in the same way as in Experiment 1. After a 12-hour break containing either sleep or wake, participants then returned for the second session. In the second session, participants again first completed the sleep habits questionnaire to ensure that they had slept for at least six hours (sleep group) or not napped throughout the day (wake group). Previously seen, unseen related, and unseen unrelated words were then presented to the centre of the screen, exactly as in Experiment 1, and participants rated their confidence of previously seeing the words, again in the same way as in Experiment 1.

2.2.1.4 Data Analysis

As in Experiment 1, we performed a series of Linear Mixed-Effects Models (Baayen, 2008; Jaeger, 2008) in R to model response accuracy and response time (RT). We again took a hypothesis-driven approach to model building, such that a model was initially fitted specifying the random effects of participants and items *(words)*. We then considered the effect of word type (old/lure/new), hemisphere (left/right), and group (sleep/wake) on model fit, before adding the two-way interaction effects and finally the three-way interaction. We ran likelihood ratio test comparisons to compare models, to indicate whether the inclusion of each factor and the interaction terms significantly improved model fit. We then assessed whether the inclusion of random effects of participants and items in the model were justified. We also conducted a Bayesian analysis on the accuracy data, to analyse the likelihood that the results provide strong evidence for the null hypothesis (H0) or the alternative hypothesis (H1).

In the same way as in Experiment 1, we also analysed both eye tracking data and confidence ratings; both overall confidence ratings for responses to old and lure words, and confidence ratings for old and lure words responded to as old. We again fitted a model with random effects of participants and items, and then built up the model by first adding main effects of word type, hemisphere, group, and eye-tracking score, and then adding the interaction terms.

2.2.2 Results

2.2.2.1 Accuracy

We again performed a series of General Linear Mixed-Effects Models, starting with a model including only random effects of participants and items, and adding in main effects of word type, hemisphere, and sleep/wake group as main effects and interaction terms. See Figure 4 for mean accuracy scores across conditions. Adding a fixed effect of word type to a model with only random effects significantly improved model fit, ($\chi^2(2) = 53.684$, $p = 2.201 \times 10^{-12}$). Both old and new words were responded to with significantly greater accuracy than lure words, replicating the same finding as in Experiment 1. Adding a main effect of hemisphere did not improve model fit compared to a model with random effects plus the main effect of word type ($\chi^2(1) = .019$, $p = .889$), nor did adding the main effect of group ($\chi^2(1) = .241$, $p = .624$). The analysis further revealed no significant interactions between word type and hemisphere ($p = .953$), word type and group ($p = .597$), hemisphere and group ($p = .967$), and no three-way interaction between word type, hemisphere, and group was evident ($p = .773$).

In the same way as for Experiment 1, we then analysed whether the inclusion of random effects was justified in the final model. Likelihood ratio tests with the same fixed effect of word type but varying in random effects indicated that the inclusion of random effect of items on intercepts was justified. There was a significant difference between a model including random effects of both participants and items on intercepts and that including only the participant effect, ($\chi^2(1) = 51.135$, $p = 8.624 \times 10^{-13}$). However, the inclusion of the random effect of participants on intercepts was not justified. There was no significant difference between a model including random effects of both participants and items on intercepts, and one including only the item effect, ($\chi^2(1) = 0$, $p = 1$).

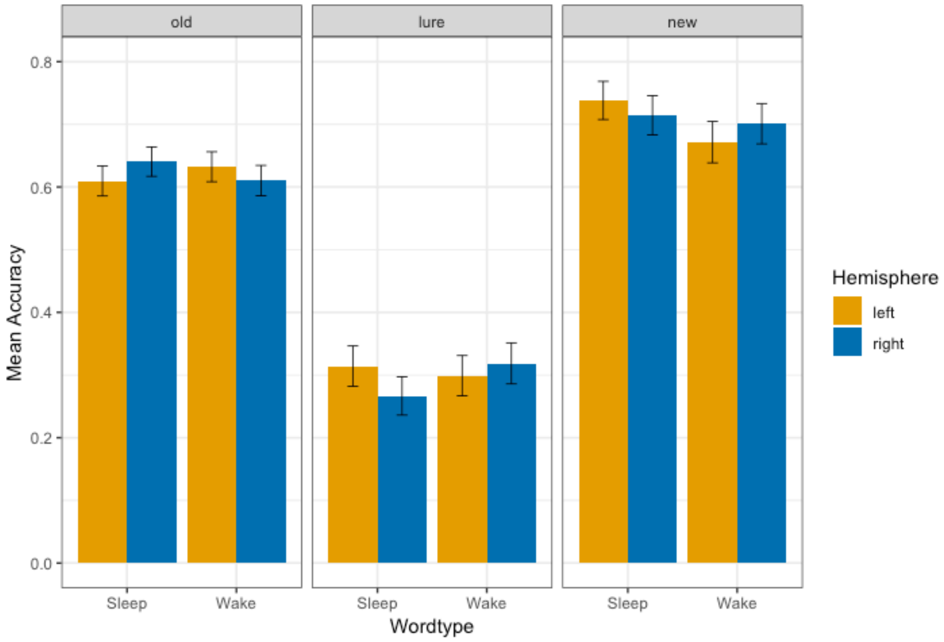


Figure 3.4. Mean response accuracy for old, lure, and new words presented to the left and right hemisphere in the sleep and wake group, error bars show ±1 SEM.

Since we found no effect of hemisphere or sleep versus wake group on performance for either previously seen old words, or unseen, related lure words, we again also ran Bayesian analyses, to assess the likelihood that the null results found here are in favour of the null (H0) hypothesis, rather than the alternative (H1) hypothesis. We used the results from Monaghan et al., (2017) to set our prior assumptions. For previously seen old words, Monaghan et al. found a significant effect of sleep, with sleep leading to increased accurate recognition specifically for words presented in the RVF/LH. Therefore, a Bayes factor greater than 1 would suggest support for this hypothesis, whereas a Bayes factor of below 1 would indicate no hemispheric asymmetry in veridical memory performance as a result of sleep. Our results revealed support for the null hypothesis, $B_{\mathrm{H}(0.07,\ 0.035)} = 0.142$. A Bayes factor of < 0.33 is considered strong evidence for the null hypothesis, thus these results suggest that when words are lateralised at encoding, rather than at retrieval, sleep does not differentially affect veridical memory performance in the LH versus RH.

For unseen, related lure words, Monaghan et al., (2017) found no significant difference in accurate rejection between the RVF/LH and LVF/RH after sleep, with sleep resulting in an overall increase in false recognition. Thus, a Bayes factor greater than 1 in this study would support the hypothesis that false memories due to sleep are not lateralised to either the LH or RH, but rather that sleep boosts spreading activation equally in both hemispheres. The current study found no difference in false recognition between the sleep and wake group, thus as expected, our results provide no strong support for either hypothesis, $\mathrm{B}_{\mathrm{H}(0.08,\ 0.04)} = 1.426$, with Bayes factors between 0.33 and 3 providing only weak evidence for either hypothesis. This suggests that there is no differential effect of sleep versus wake on false recognition when word lists are lateralised at encoding.

For old and lure words, we again added the factor of eye tracking score to the model, to establish whether participants' fixation influenced the lack of a hemispheric effect. See Appendix I for mean eyetracking score across conditions. Adding the main effect of eye tracking score to a model with random effects and a main effect of word type did not improve model fit ($\chi^2(1) = .022$, $p = .881$), nor did adding the interaction terms word type by eye tracking score ($p = .524$), hemisphere by eye tracking score ($p = .908$) or group by eye tracking score ($p = .998$). There was also no significant three-way interaction between word type, group, and eye tracking score ($p = .941$), word type, hemisphere, and eye tracking score ($p = .844$), group, hemisphere, and eye-tracking score ($p = .998$), and no four-way interaction ($p = .978$).

2.2.2.2 Confidence Ratings

We then analysed whether there was a difference in confidence ratings for old and lure words, following the same analysis as in Experiment 1. See Figure 5 for mean confidence rating across conditions. Again, there was a main effect of word type ($\chi^2(1) = 5.578$, $p = .018$), with lure words being rated as previously seen more confidently than old words (estimate = .359, SE = .146, $t = 2.457$). No other main effects or interactions were significant ($p > .05$).

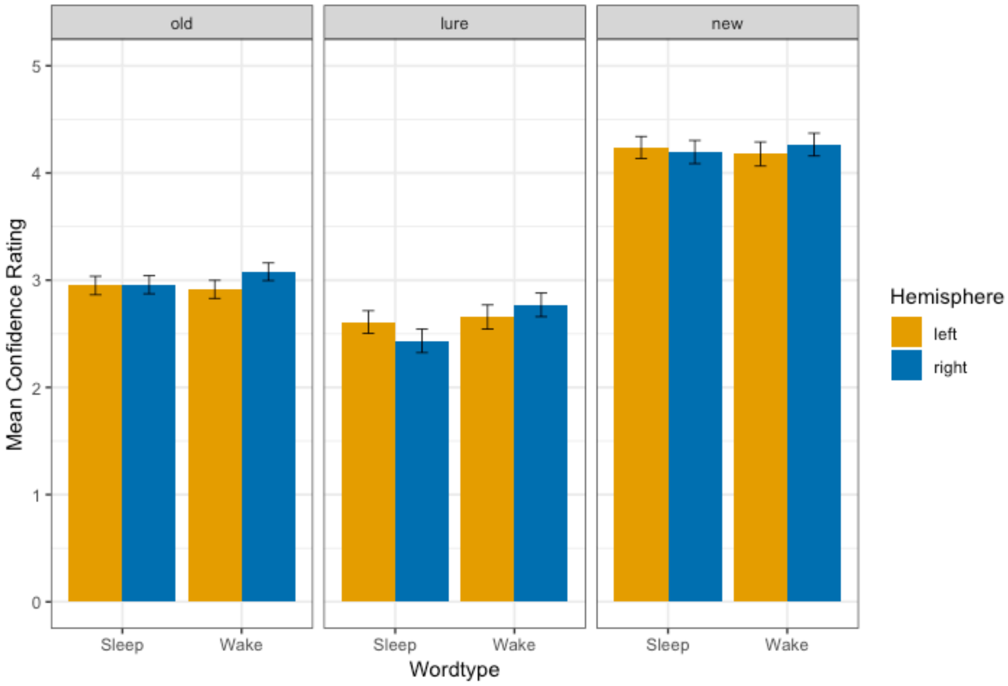


Figure 3.5. Mean confidence ratings for old, lure, and new words presented to the left and right hemisphere in the sleep and wake groups, error bars show ±1 SEM.

We also analysed confidence ratings specifically for old and lure words that were identified as old. Adding word type to a model with only random effects did not improve model fit, ($\chi^2(1) = 1.280$, $p = .258$), indicating no significant difference in confidence in accepting old or lure words as previously seen. Adding hemisphere to a model with only random effects of participants and items did marginally improve model fit, ($\chi^2(1) = 3.403$, $p = .065$). Participants rated words more confidently as old if the list was encoded in the RVF/LH (estimate = .072, SE = .039, $t = 1.850$). The analysis also revealed a significant interaction between hemisphere and eye tracking scores ($\chi^2(3) = 12.674$, $p = .005$). For word lists encoded in the LVF/RH, participants were much less confident in responding 'old' if they had a lower eye tracking score (Figure 6) compared with those words with word lists encoded in the RVF/LH with a low eye tracking score (Figure 7).

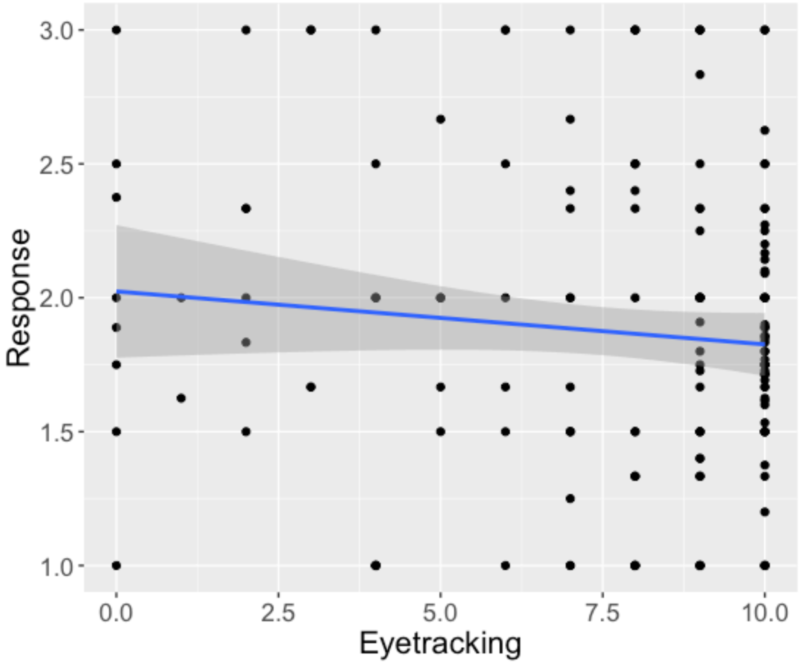


Figure 3.6. The effect of eye tracking scores on confidence ratings for old and lure words rated as old in the RH.

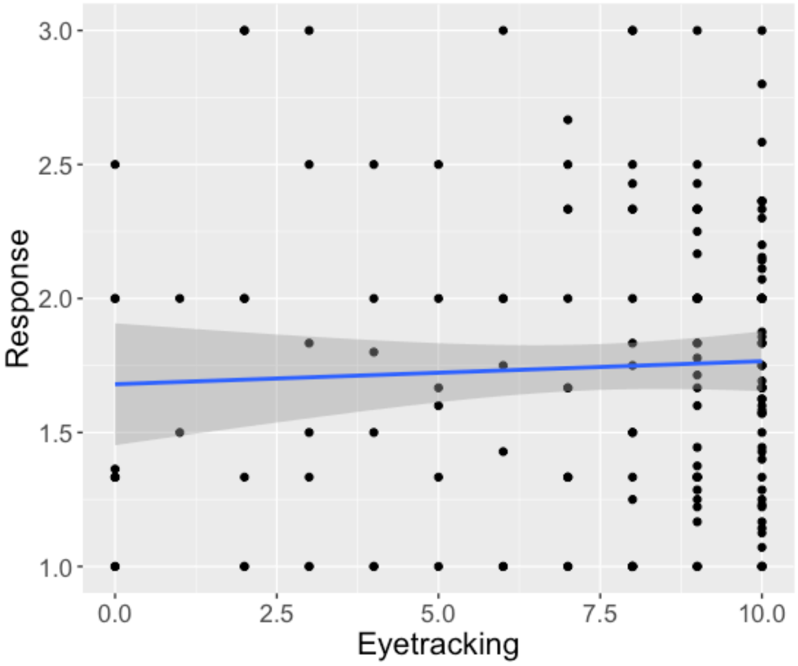


Figure 3.7. The effect of eye tracking scores on confidence ratings for old and lure words rated as old in the LH.

2.2.2.3 Response Times

For the RT analysis of correct responses, response times exceeding 2.5 standard deviations from the mean were removed from the analysis, leading to the removal of 2.136% of the data, and leaving 1,826 data points. See Figure 8 for mean response time across conditions. Again, we found that adding word type to a model with only random effects of participants and items significantly improved model fit, ($\chi^2(2) = 24.849$, $p = 4.018 \times 10^{-6}$). New words were accurately rejected significantly faster than lure words, and old words were accurately accepted quicker than lure words were rejected.

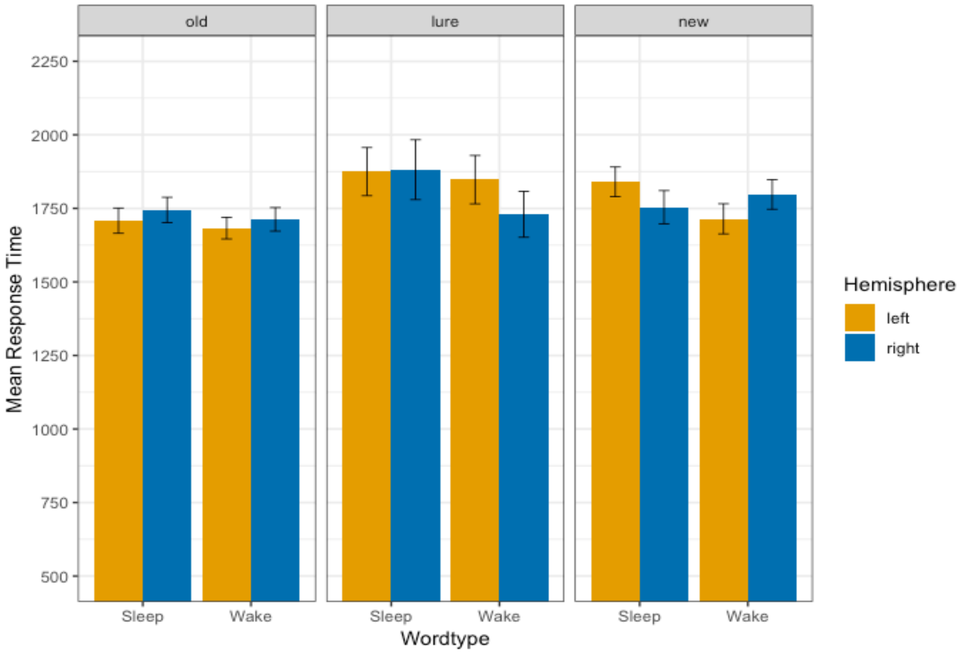


Figure 3.8. Mean response times for old, lure, and new words presented to the left and right hemisphere in the sleep and wake groups, error bars show ±1 SEM.

Adding the 2-way interaction between word type and group also improved model fit, ($\chi^2(3) = 8.670$, $p = .034$) (See Figure 9). To investigate this interaction further, we first split the data between word types. For old and lure words, adding the main effect of group did not improve model fit. However, for new words there was a significant effect of group ($\chi^2(1) = 5.229$, $p = .022$). The sleep group were slower to accurately reject new words than the wake group, (estimate = -.085, SE = .036, t = -2.327). We then analysed the effect of word type on the sleep and wake group separately. For the wake group, adding the main effect of word type to a model with only random effects improved model fit, ($\chi^2(2) = 19.876$, $p = 4.83 \times 10^{-5}$). As expected, old and new words were responded to quicker than lure words (estimate = -.092, SE = .020, t = -4.655; estimate = -.086, SE = .021, t = -4.034). There was no significant difference in RT between old and new words (estimate = -.005, SE = .016, t = -.321). There was also a main effect of word type for the sleep group, ($\chi^2(2) = 13.537$, $p = .001$). Old words were responded to quicker than both lure words (estimate = -.070, SE = .021, t = -3.272) and new words (estimate = -.046, SE = .016, t = -2.910). There was no significant difference between lure and new words, (estimate = -.024, SE = .023, t = -1.043)

For old and lure words, we again analysed whether eye tracking scores had any effect on the results. We found that adding the main effect of eye tracking score to a model with random effects and the main effect of word type did not significantly improve model fit, nor did adding any interaction terms (for all results, $p > .05$).

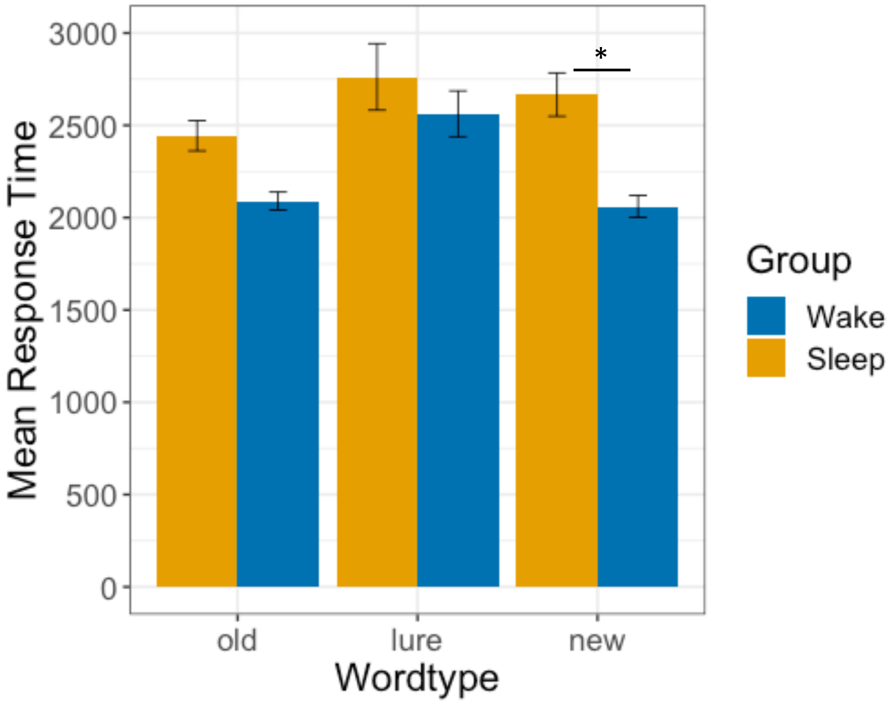


Figure 3.9. Response times for accurate responses to old, lure, and new words in the sleep and wake groups separately, error bars show ±1 SEM, **p* < .05.

3. Discussion

A main aim of the study was to investigate whether hemispheric asymmetries in veridical and false memories are evident when word lists are lateralised at encoding, and what role sleep plays in the processing of veridical and false memories in the RVF/LH and LVF/RH. If encoding word lists in either the LH or RH has a similar effect to lateralisation effects found during retrieval (Bellamy & Shillcock, 2007; Faust et al., 2008; Ito, 2001; Westerberg & Marsolek, 2003), then we would expect to find an increase in veridical memory for previously seen words that were presented to the LH during encoding, and a RH bias for false recognition of unseen, related lure words. However, evidence suggests that there may not be any LH versus RH differences in memory performance at encoding (Fabiani et al., 2000; Hocking et al., 2009). The current results support the latter, that encoding word lists in the LH or RH has no significant effect on either veridical or false memories. A Bayesian analysis based on previous findings from Bellamy and Shillcock (2007), provides strong support for this, again indicating no effect of lateralisation at encoding on either veridical or false memories. Thus, the LH advantage for veridical memory performance and RH bias for false memories is the consequence of reactivation and retrieval of those memories from either the fine semantic associations of the LH, or the broad semantic associations of the RH.

A second main aim of the present study was to examine whether sleep differentially effects memory performance when word lists are encoded in the LH versus RH. Previous research by Monaghan et al., (2017) found that when words are presented centrally during encoding, and lateralised during recognition, then there is a shift to a LH advantage for veridical memory after sleep. This supports literature indicating a reactivation of newly acquired words during consolidation from hippocampal systems to long-term representations in the LH (Diekelmann & Born, 2010). The present study did not replicate these findings, revealing no difference in veridical memory performance after sleep between words taken from word lists encoded in either the LH or RH. Whether word lists where encoded in the RH or LH therefore has no effect on how they are subsequently reactivated, consolidated, and retrieved.

The second experiment also investigated whether there were any differential effects of sleep versus wake on false memories when encoding word lists in either hemisphere. Monaghan et al. (2017) suggest that when words are lateralised at retrieval, sleep leads to spreading activation equally across the two hemispheres, with an overall increase in false recognition after sleep, rather than a specific hemispheric effect. This was not replicated by the current study; no significant difference in false recognition was evident between the sleep and wake group, and no lateralisation effects were found. Thus, encoding word lists in the LH versus RH has no effect on how the information is reactivated and consolidated for later retrieval, and instead we suggest that word lists are activated equally across hemispheres.

One possible explanation for this lack of an overall effect of sleep on false memories is due to the use of recognition testing, rather than recall. Previous research (Diekelmann, Landolt, Lahl, Born, & Wagner, 2008; Fenn, Gallo, Margoliash, Roediger, & Nusbaum, 2009) and a recent meta-analysis (Newbury & Monaghan, 2018), indicate no effect of sleep on false memories when a recognition task is used, whereas if participants are given tests of recall, then sleep leads to increases in false memories. The activation/monitoring framework (Collins & Loftus, 1975) provides an explanation as to why differences between tests of recall and recognition occur in DRM tests. When words are presented to participants during a recognition task, it is easier to accurately reject unseen, related words, due to monitoring cues that are activated (Watson, McDermott & Balotta, 2004). These source monitoring abilities are further enhanced after sleep (Johnson, Hashtroudi, & Lindsay, 1993). During tests of free recall, participants must retrieve words from their stored memory, thus monitoring cues are not available, and unseen, related words are more likely to be activated and falsely retrieved.

Experiment 2 also indicated a marginal effect of hemisphere on confidence ratings for old and lure words responded to as old. The results suggest that words were less confidently rated as old when the DRM list associated with that word was encoded in the RH than the LH. This supports theories of RH broad semantic coding

(Beeman & Bowden, 2000; Beeman & Chiarello, 1998; Burgess & Simpson, 1988; Monaghan et al., 2004), as participants may be less likely to distinguish previously seen versus unseen, related words, and thus report an overall lower confidence in their decisions. There was also an interaction between hemisphere and eye tracking score that suggests confidence ratings in the RH were lower when participants also had a lower eye tracking score. There was no significant difference in confidence ratings between the LH and RH when eye tracking scores were greater. Higher eye tracking scores mean that we can be sure that participants were fixating on the centre of the screen, thus word lists were being encoded in only the RVF/LH or LVF/RH. Since we find no other evidence of hemispheric asymmetry when word lists are lateralised at encoding, we cannot make conclusions regarding the effect of lateralised encoding of word lists on subsequent confidence ratings.

The two experiments outlined here examine the effect of lateralisation at encoding on subsequent veridical and false memories, and whether sleep plays any role in the processing of these memories. The findings suggest that whether word lists are encoded in the RH or LH has no effect on overall memory performance. This is in contrast to studies suggesting an effect of lateralisation at retrieval that indicate a LH advantage for veridical memory and a RH bias for false memories (Bellamy & Shillcock, 2007; Faust et al., 2008; Ito, 2001; Westerberg & Marsolek, 2003). Further, we found no effect of sleep on either veridical or false memories, supporting evidence for a lack of an effect of sleep on recognition testing (Newbury & Monaghan, 2018). Whereas Monaghan et al., (2017) suggest a boost to LH veridical memory after sleep, in line with previous research suggesting reactivation of memories into the neo-cortical systems in the LH, the present study suggests no differential effects of sleep on lateralisation at encoding. Thus, we conclude that hemispheric processing of memories and subsequent effects of sleep due to LH fine semantic coding and RH broad semantic coding are specific to retrieval processes.

References

Baayen, R. H. (2008). *Analyzing linguistic data [electronic resource] : a practical introduction to statistics using R*. Cambridge: Cambridge : Cambridge University Press.

Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological Influences on the Recognition of Monosyllabic Monomorphemic Words. *Journal of Memory and Language, 55*(2), 290-313. doi:10.1016/j.jml.2006.03.008

Beeman, M. J., & Bowden, E. M. (2000). The right hemisphere maintains solution-related activation for yet-to-be-solved problems. *Memory & Cognition, 28*(7), 1231-1241. doi:10.3758/BF03211823

Beeman, M. J., & Chiarello, C. (1998). Complementary right-and left-hemisphere language comprehension. *Current Directions in Psychological Science, 7*(1), 2-8.

Beeman, M. J., Friedman, R. B., Grafman, J., Perez, E., Diamond, S., & Lindsay, M. B. (1994). Summation priming and coarse semantic coding in the right hemisphere. *Journal of Cognitive Neuroscience, 6*(1), 26-45. doi:10.1162/jocn.1994.6.1.26

Bellamy, K. J., & Shillcock, R. (2007). A right hemisphere bias towards false memory. *Laterality, 12*(2), 154-166. doi:10.1080/13576500601051648

Bolduc, C., Daoust, A., Limoges, É., Braun, C. M. J., & Godbout, R. (2003). Hemispheric lateralization of the EEG during wakefulness and REM sleep in young healthy adults. *Brain and Cognition, 53*(2), 193-196. doi:10.1016/S0278-2626(03)00108-8

Bui, D. C., Friedman, M. C., McDonough, I. M., & Castel, A. D. (2013). False memory and importance: Can we prioritize encoding without consequence?. *Memory & cognition*, *41*(7), 1012-1020.

Burgess, C., & Simpson, G. B. (1988). Cerebral hemispheric mechanisms in the retrieval of ambiguous word meanings. *Brain and Language, 33*(1), 86-103. doi:10.1016/0093-934X(88)90056-9

Cai, D., Mednick, S. A., Harrison, E, M., Kanady, J. C., & Mednick, S. C. (2009). REM, not incubation, improves creativity by priming associative networks. *Proceedings of the National Academy of Sciences, 106*, 10130–10134. doi:10.1073/pnas.0900271106

Casagrande, M., & Bertini, M. (2008). Night-time right hemisphere superiority and daytime left hemisphere superiority: A repatterning of laterality across wake–sleep–wake states. *Biological Psychology, 77*(3), 337-342. doi:10.1016/j.biopsycho.2007.11.007

Casagrande, M., Violani, C., De Gennaro, L., Braibanti, P., & Bertini, M. (1995). Which hemisphere falls asleep first? *Neuropsychologia, 33*(7), 815-822. doi:10.1016/0028-3932(95)00004-M

Chiarello, C., Burgess, C., Richards, L., & Pollock, A. (1990). Semantic and associative priming in the cerebral hemispheres: Some words do, some words don't … sometimes, some places. *Brain and Language, 38*(1), 75-104. doi:10.1016/0093-934X(90)90103-N

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407-428. doi:10.1037/0033-295X.82.6.407

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience, 11*(2), 114-126. doi:10.1038/nrn2762

Diekelmann, S., Born, J., & Wagner, U. (2010). Sleep enhances false memories depending on general memory performance. *Behavioural Brain Research, 208*(2), 425-429. doi:10.1016/j.bbr.2009.12.021

Diekelmann, S., Landolt, H., Lahl, O., Born, J., & Wagner, U. (2008). Sleep loss produces false memories. *PloS one, 3*(10), e3512-e3512. doi:10.1371/journal.pone.0003512

Dienes, Z., & McLatchie, N. (2017). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin & Review, 25*(1), 207-218. doi: 10.3758/s13423-017-1266-z.

Ellis, A. (2004). Length, formats, neighbours, hemispheres, and the processing of words presented laterally or at fixation. *Brain and Language, 88*(3), 355-366. doi:10.1016/S0093-934X(03)00166-4

Fabiani, M., Stadler, M. A., & Wessels, P. M. (2000). True But Not False Memories Produce a Sensory Signature in Human Lateralized Brain Potentials. *Journal of Cognitive Neuroscience, 12*(6), 941-949. doi:10.1162/08989290051137486

Faust, M., Ben-Artzi, E., & Harel, I. (2008). Hemispheric asymmetries in semantic processing: Evidence from false memories for ambiguous words. *Brain and Language, 105*, 220–228.

Fenn, K. M., Gallo, D. A., Margoliash, D., Roediger, H. L., & Nusbaum, H. C. (2009). Reduced false memory after sleep. *Learning & Memory, 16*(9), 509-513. doi: 10.1101/lm.1500808

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *38*(7), 833-848.

Hocking, J., McMahon, K. L., & De Zubicaray, G. I. (2009). Semantic context and visual feature effects in object naming: An fMRI study using arterial spin Labeling. *Journal of Cognitive Neuroscience,21*(8), 1571-1583. doi:10.1162/jocn.2009.21114

Ito, Y. (2001). Hemispheric asymmetry in the induction of false memories. *Laterality, 6*(4), 337-346. doi:10.1080/13576500143000050

Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (Transformation or Not) and towards Logit Mixed Models. *Journal of Memory and Language, 59*(4), 434-446. doi:10.1016/j.jml.2007.11.007

Jeffreys, H. (1939/1961). *The theory of probability (1st/3rd Ed.)*. Oxford: Oxford University Press.

Johnson, M., Hashtroudi, S., Lindsay, D., & Steinberg, R. J. (1993). Source Monitoring. *Psychological Bulletin, 114*(1), 3-28. doi: 10.1037/0033-2909.114.1.3

Kim, H., & Cabeza, R. (2006). Differential contributions of prefrontal, medial temporal, and sensory-perceptual regions to true and false memory formation. *Cerebral Cortex*, *17*(9), 2143-2150.

Kim, H., & Cabeza, R. (2006). Differential contributions of prefrontal, medial temporal, and sensory-perceptual regions to true and false memory formation. *Cerebral Cortex*, *17*(9), 2143-2150.

Marshall, L., & Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends in Cognitive Sciences, 11*(10), 442-450. doi:10.1016/j.tics.2007.09.001

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*(3), 419–457. doi:10.1037/0033-295X.102.3.419

Monaghan, P., Shaw, J. J., Ashworth-Lord, A., & Newbury, C. R. (2017). Hemispheric processing of memory is affected by sleep. *Brain and Language, 167,* 36-43. doi:10.1016/j.bandl.2016.05.003

Monaghan, P., Shillcock, R., & McDonald, S. (2004). Hemispheric asymmetries in the split-fovea model of semantic processing. *Brain and Languag., 88*(3), 339-354. doi:10.1016/S0093-934X(03)00165-2

Newbury, C. R., & Monaghan, P. (2018). When does sleep affect veridical and false memory consolidation?: A meta-analysis. *Psychonomic Bulletin & Review*, 1-14.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113.

Payne, J. D., Schacter, D. L., Propper, R. E., Huang, L., Wamsley, E. J., Tucker, M. A., . . . Stickgold, R. (2009). The role of sleep in false memory formation. *Neurobiology of Learning and Memory, 92*(3), 327-334. doi:doi:10.1016/j.nlm.2009.03.007

Pirozzolo, F. J., & Rayner, K. (1977). Hemispheric specialization in reading and word recognition. *Brain and Language, 4*(2), 248-261. doi:10.1016/0093-934X(77)90021-9

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rasch, B., & Born, J. (2007). Maintaining memories by reactivation. *Current Opinion in Neurobiology, 17*(6), 698-703. doi:10.1016/j.conb.2007.11.007

Shaw, J. J., & Monaghan, P. (2017). Lateralised sleep spindles relate to false memory generation. *Neuropsychologia, 107*, 60-67. doi:10.1016/j.neuropsychologia.2017.11.002

Sio, U. N., Monaghan, P., & Ormerod, T. (2013). Sleep on it, but only if it is difficult: Effects of sleep on problem solving. *Memory and Cognition*, *41*, 159-166. doi:10.3758/s13421-012-0256-7

Stickgold, & Walker. (2005). Memory consolidation and reconsolidation: What is the role of sleep? *Trends in Neurosciences, 28*(8), 408-415. doi:10.1016/j.tins.2005.06.004

Straube, B. (2012). An overview of the neuro-cognitive processes involved in the encoding, consolidation, and retrieval of true and false memories. *Behavioral and Brain Functions*, *8*(1), 35.

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *Journal of Neuroscience*, *30*(43), 14356-14360.

Walker, M. P., & Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron, 44*(1), 121-133. doi:10.1016/j.neuron.2004.08.031

Watson, J. M., McDermott, K. B., & Balota, D. A. (2004). Attempting to avoid false memories in the Deese/Roediger-McDermott paradigm: Assessing the combined influence of practice and warnings in young and old adults. *Memory & Cognition, 32*(1), 135-141. doi:10.3758/BF03195826.

Westerberg, C. E., & Marsolek, C. J. (2003). Hemisphere asymmetries in memory processes as measured in a false recognition paradigm. *Cortex, 39*(4-5), 627-642.

4. Appendix

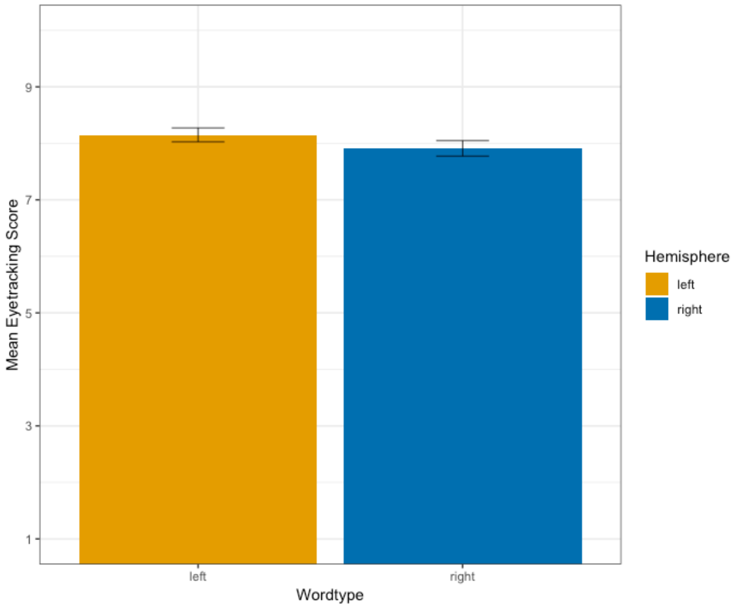


Figure 3.A1. Mean eye tracking score for words presented to the left and right hemisphere.

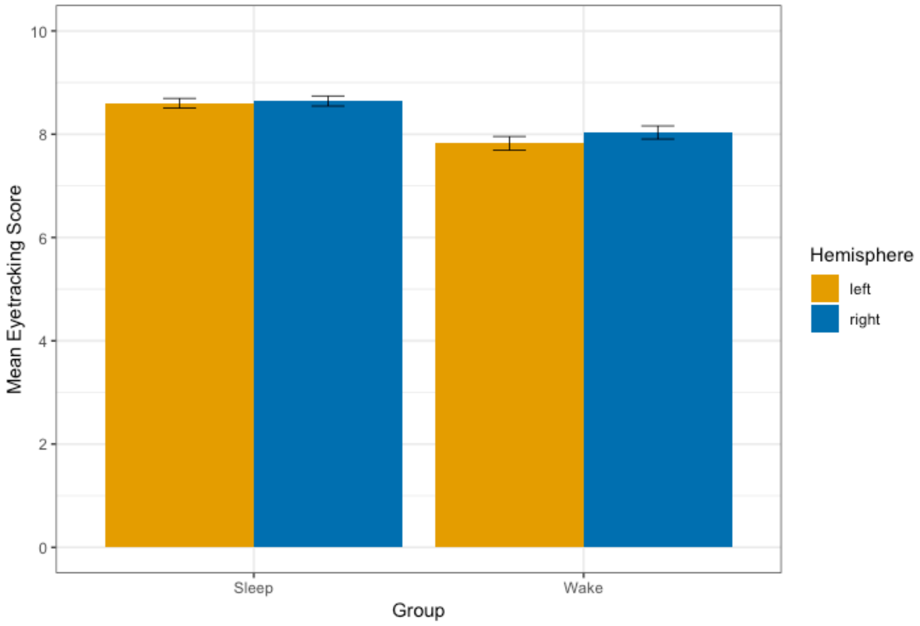


Figure 3.A2. Mean eye tracking score for words presented to the left and right hemisphere in the sleep and wake group.

Chapter 4: **Negative but not Positive Emotional Memories are Enhanced by Sleep.**

The previous chapter demonstrated that effects of lateralisation on both veridical and false memories found at recognition does not extend to encoding processes, in both an immediate testing task and after a delay including overnight sleep or daytime wakefulness. Another factor that has been found to influence both veridical and false memories is that of emotionality of the stimuli. Previous findings suggest that emotional memories are more accurately remembered than neutral stimuli, and there are suggestions of an effect of emotionality on the formation of false memories. Furthermore, similarly to lateralisation effects, sleep has been suggested to boost memory for emotional stimuli. The following chapter seeks to replicate findings of greater memory performance for emotional veridical memories, and also explore whether this emotionality effect extends to the formation of false memories.

Statement of Author Contribution.

In the chapter entitled “Negative but not Positive Emotional Memories are Enhanced by Sleep”, the authors agree to the following contributions:

Chloe R. Newbury –80% (Experimental design, data collection, data analysis, writing)

Signed: C.Newbury Date: 25/04/2019

Professor Padraic Monaghan – 20% (Experimental design and review)

Signed: P Mon Date: 25/04/2019

A subset of the data in Experiment 1 was collected and analysed as part of my MSc in Psychological Research Methods at Lancaster University, 2015.

Abstract

Memory for emotional stimuli are enhanced compared to neutral stimuli, and negative emotions appear to be particularly supported by consolidation during sleep compared to wake. However, previous studies of sleep and emotional memory have not directly compared positive and negative emotional stimuli, nor have they examined the relative effect of sleep and emotion on the formation of veridical and false memories. In Experiment 1, participants were tested on recognition performance of positive and negative emotional and neutral lists of thematically-related words, after a 12-hour period containing sleep or wake. In Experiment 2 participants were tested with no delay to test for time-of-day effects. We found that sleep particularly enhanced veridical memory of negative compared to neutral words, whereas wake led to more veridical and false memories for positive words. The results suggest that wake consolidates positive memories, whereas sleep results in enhanced memory for negative information.

1. Introduction

Emotion affects the encoding and processing of memories, and it is well-known that memory for emotional events or information is more accurate than memory for neutral information (Adelman & Estes, 2013; Kensinger & Corkin, 2004a; McGaugh, 2004). These effects seem to be observed regardless of whether emotion is positive or negative. In studies of emotional word processing, words of negative valence (Kensinger & Corkin, 2003), and positive valence (Adelmann & Estes, 2013) lead to more accurate memory performance on subsequent tests than neutral words. Further, words rated as high in arousal have also been found to lead to increased memory performance compared to neutral words, to a greater extent than the effects of valence (Kensinger & Corkin, 2003).

The effects of emotional content on the generation of false memories – memory for information that was not experienced – is less clear. A classic test of the mechanisms and extent to which false memories are generated uses the Deese-Roediger-McDermott (DRM) paradigm (Roediger & McDermott, 1995), where participants are exposed to lists of semantically related words (eg., *bed, dream, tired, snooze, yawn,* etc.). Participants are then asked to recall or identify words that they had previously seen in the initial lists. Words are categorised as either those that appeared in the initial lists (old words), words that did not appear in the lists but were closely related, known as unseen, related lure words (eg., *sleep* in the above list), or unseen, unrelated words (new words). Participants are more likely to recall, or identify as previously seen, lure words than new words, indicating a false memory effect (McDermott, 1996; Roediger & McDermott, 1995; Roediger, Watson, McDermott, & Gallo, 2001).

Pesta, Murphy, and Sanders (2001) found that participants were more likely to accurately reject emotional lure words that were orthographically associated to DRM word lists compared to neutral lure words, suggesting that emotionality reduces the false recognition effect (see also Kensinger & Corkin 2004b). However, they also found that adding three more emotive words to the initial lists during learning, thereby decreasing the distinctiveness of individual emotional words, increased false recognition of emotional lure words.

More recent research has thus aimed to control for distinctiveness of emotional word lists to investigate the specific effect of emotionality on false memories. These studies have found an increase, rather than a decrease, in the false recognition of emotionally negative word lists (El Sharkawy, Groth, Vetter, Beraldi, & Fast, 2008; Howe, Candel, Otgaar, Malone, & Wimmer, 2010). Furthermore, Brainerd, Stein, Silveira, Rohenkohl, and Reyna (2008), who aimed to develop emotive DRM word lists that controlled for distinctiveness and arousal, as well as word frequency, concreteness and backwards associative strength (BAS), found a significant difference in the processing of negative and positive word lists in tests of recognition. Using emotive and neutral DRM word lists, they found an increase in false recognition of negative emotional lure words, supporting previous findings (El Sharkawy et al., 2008; Howe et al., 2010). However positive word lists elicited a decrease in false recognition compared to neutral lists, suggesting that there may be differences in the processing of emotionally negative and positive information.

Studies of the effect of recall, rather than recognition, on emotional false memories have found mixed results. Bauer, Olheiser, Altarriba and Landi (2009) report a significant effect of emotionality on false recall performance; emotional words were falsely recalled to a greater extent than both concrete and abstract neutral words, with false recall of positive words driving the effect. On the other hand, Howe et al. (2010) found that neutral DRM lists elicited increased false recall compared with negative lists, contrasting with their finding that false recognition increases when participants are exposed to emotional than neutral lists. This indicates that there may be differences in the processing of emotionally positive and negative word lists during tests of recognition compared to tests of recall.

A range of evidence indicates that memories are affected by the offline processes involved in sleep, including encoding, consolidation and retrieval of both declarative and procedural memories (Rasch & Born, 2007, 2013; Stickgold, 2005). Memory consolidation during sleep leads to the generalisation and strengthening of selective memories. According to the Active Systems Consolidation hypothesis (Diekelmann & Born, 2010; Marshall & Born, 2007; Rasch & Born, 2013) this is due to memories initially encoded in the hippocampus and neocortical systems during wakefulness being selectively reactivated and stored long-term in the neocortex. Sleep may also result in increased activation in long-term semantic associative memory, which can strengthen memory traces both for experienced stimuli, as well as increase activation of related but unseen information, as in the DRM lure words (Cai, Mednick, Harrison, Kanady, & Mednick, 2009; Monaghan, Shaw, Ashworth-Lord, & Newbury, 2017; Sio, Monaghan, & Ormerod, 2013).

Emotional information in particular has been found to benefit from the positive effects of sleep. Payne, Stickgold, Swanberg, and Kensinger (2008) found that participants who slept between intial exposure and later testing were significantly more likely to remember the emotional object of a scene in comparison to the neutral components of the scene. This improvement in memory for emotional content did not extend to those participants in the wake condition. Research has also found a similar effect when investigating memory performance for positive versus neutral stimuli (Chambers & Payne, 2014), with those who slept prior to testing showing an increase in memory for humorous cartoons compared to non-humorous cartoons. Relatedly, Hu, Stylos-Allan, and Walker (2006) found that more arousing stimuli were more accurately recognised after a period of sleep compared to wake, and Nishida, Pearsall, Buckner, and Walker (2009) found that participants who napped between exposure and subsequent testing showed increased consolidation of emotional content compared to neutral information, which was related to the duration of rapid eye movement (REM) sleep a participant had.

A number of studies indicate a significant role of REM sleep in the processing of emotional memories (Carr & Nielsen, 2015; Goldstein & Walker, 2014). It is suggested that REM sleep supports emotional memory performance due to specific characteristics that occur during REM sleep, including increased levels of limbic and forebrain acetylcholine (Vazquez & Baghdoyan, 2001), found to be important in the development of emotional memories (McGaugh, 2004). Carr and Nielsen (2005) hypothesise that this REM sleep advantage leads to a greater spread of activation of emotional compared to neutral stimuli. Although this allows for quicker and easier access to emotional memories, sleep-induced spreading activation also enhances the likelihood of producing unseen, related information, due to increased activation of more semantic associations (Howe, Wimmer, Gagnon, & Plumpton, 2009; Roediger et al., 2001). Walker and Stickgold (2004) and van der Helm and Walker (2009) further suggest a role of both slow wave sleep (SWS) and REM sleep in the consolidation of emotionally salient information and events. These memories are particularly strongly consolidated into the long-term store in terms of the information content of the memory, but with a consequent reduction in the strength of the emotion attached to the event or experience, suggesting an adaptive nature of memory consolidation.

Emotional information has been proposed to be susceptible to greater spreading activation in semantic associative memory, which can then result in increased false memories due to activation of related but unseen information within the semantic associative network (Carr & Nielsen, 2005). If sleep increases activation in semantic memory (Cai et al., 2009) then this could also potentially lead to greater enhancement of emotional false memories following sleep. However, despite the overlap of potential mechanisms, only one study to our knowledge has investigated the link between emotionality, sleep, and the production of false memories. McKeon, Pace-Schott, and Spencer (2012), using negative emotional and neutral DRM word lists, found that sleep increased veridical memory for neutral but not negative word lists, conflicting with previous research showing that sleep enhanced memory for emotional stimuli (Hu et al., 2006; Nishida et al., 2009; Payne et al., 2008). McKeon et al., (2012) also found that sleep led to an increase in false recall of both negative and neutral lure words, supporting previous research to suggest that sleep leads to the consolidation of a general gist or broader spread of activation (Diekelmann & Born, 2010; Payne et al., 2009; Straube, 2012). They also found that neutral lure words were falsely recalled significantly more than negative lure words for both wake and sleep groups, consistent with previous research (Howe et al., 2010). However, McKeon et al. did not control for effects of arousal, thus effects of sleep resulting from comparisons between negative and neutral word lists could be a consequence of differences in valence or differences in arousal between these words, which can affect veridical memory (Kensinger & Corkin, 2003).

Furthermore, there have been few studies investigating effects of sleep on positive emotional stimuli, and none using the DRM paradigm.

For neutral DRM word lists, there are differences in the effect of sleep on tests of recall compared to recognition. Payne et al. (2009) found that participants who slept between first exposure to DRM word lists and subsequent testing correctly recalled significantly more old words, but also falsely recalled more lure words than those participants who stayed awake (see also Diekelmann, Born & Wagner, 2010). In the same set of studies, Payne et al., also found that even a short period of sleep in the form of a daytime nap led to enhanced false recall of the lure words. The effect of sleep on false recognition is however less clear. Fenn, Gallo, Margoliash, Roediger, and Nusbaum (2009) found that those participants who slept between learning and testing of DRM word lists showed a reduction in false recognition compared to those participants who stayed awake. Furthermore, Diekelmann, Landolt, Lahl, Born, and Wagner (2008) found no significant difference in false recognition performance between participants who slept and those who stayed awake, and Monaghan et al., (2017) did find an enhanced false memory effect for a recognition test after sleep. Indeed, a recent meta-analysis conducted by Newbury and Monaghan (2018) aimed to clarify these effects, and found that sleep significantly increases false recall, but has no significant effect on false recognition for neutral word lists.

Consequently, the current study investigated whether sleep differentially affected the processing of negative, neutral *and* positive words, with each of the word lists controlling for a number of variables, including arousal. We chose to conduct a recognition memory task because previous research suggests that false recognition of emotional lure words results in a larger effect than that found for a recall task (Howe et al., 2010; Sharkawy et al., 2008). Furthermore, whereas sleep has been found to affect recall memory for negative emotional stimuli (McKeon et al., 2012), it is unclear whether emotion will affect recognition memory in the same way, given that, for neutral stimuli, sleep has been observed to have mixed effects on veridical and false memory (Diekelmann et al., 2008; Fenn et al., 2009; Monaghan et al., 2017; Shaw & Monaghan, 2017).

Based on previous research, we predicted that emotionality of the word lists would elicit increased performance accuracy for previously seen words (Adelman & Estes, 2013; Kensinger & Corkin, 2004a; McGaugh, 2004), as well as increased false recognition of emotional lure words (El Sharkawy et al., 2008; Howe et al., 2010). Furthermore, we hypothesised that sleep, in contrast to wake, would enhance accurate memory performance for previously seen negative words compared to neutral words, and increase false recognition of negative unseen, related lure words in comparison to neutral lure words. Finally, if sleep affects processing of emotional stimuli regardless of valence, then we would expect that positive words will also be more accurately recognised after sleep compared to neutral words and we would also observe increased levels of false memory for positive emotion words. However, previous studies have indicated reduced recognition of positive compared to neutral lure words for studies involving a test soon after initial exposure to DRM lists (Brainerd et al., 2008). Therefore, effects of positive compared to negative emotional words may be differentially affected by sleep. If this is the case, then models of sleep on emotional memory (e.g., van der Helm & Walker, 2009; Walker & Stickgold, 2004) will need to be tailored according to the valence of the stimuli. In Experiment 1 we tested the effect of sleep and wake on emotional and neutral veridical and false memory. In Experiment 2, we determined whether any sleep effects observed in the first experiment were due to circadian effects associated with time of day of training and testing.

2.1 Experiment 1: Effects of sleep and wake on the processing of emotional and neutral word lists

In this study, we presented participants with DRM lists containing positive or negative, and neutral words, and then, after a 12-hour interval including sleep or wake, tested participants on their recognition of previously seen words, unseen unrelated new words, and unseen related lure words. If valence, rather than arousal, affects the role of sleep for memory of words, then we would anticipate a significant interaction on memory for positive compared to negative emotional words, with greater performance for negative than positive words (Brainerd et al., 2008). However, if the effect of sleep on memory is more generally applied to any emotionally arousing stimulus, then no differences in memory for positive versus negative emotional words would be observed.

2.1.1 Methods

2.1.1.1 Participants

Sixty participants (47 female, 13 male) with a mean age of 20 years (SD = 2.88, range = 18 to 29) took part, who were either personal associates of the researcher or undergraduate students who took part in the study for course credit. They were assigned to one of four groups. Sample size was based on McKeon et al. (2012), who tested 15 participants per group, with power = .74 for the effect of sleep versus wake on emotional lure words. All participants gave informed consent and were fully debriefed at the end of the study. Prior to data collection we established exclusion criteria that led to the removal of 51 additional participants: in the same way as previous experiments investigating effects of sleep, participants in the sleep condition who slept for less than 6 hours (*n* = 22) were removed from the analyses (e.g., Diekelmann et al., 2008; McKeon et al., 2012; Payne et al., 2008), as well as those aged over 31 years old (*n* = 6). This cut-off was based on previous studies that used the same cut-off point, due to findings that adults 32 and over begin to show increasing levels of false memories (Dennis, Kim & Cabeza, 2007; Kensinger & Corkin, 2004b). Participants who were non-native English speakers (*n* = 30) were also not included in the final analyses, but took part in the experiment as part of their course credit. Note that results for all 111 participants are presented in Appendix A, and the key effects were still observed.

2.1.1.2 Materials

Negative and positive emotion DRM word lists and related lures were taken from Brainerd, Holliday, Reyna, Yang, and Toglia (2010), who controlled for properties of the word lists that are thought to affect false memory formation, including concreteness, meaningfulness, and frequency of words (Roediger et al., 2001). Neutral DRM lists and their related lures were taken from Stadler, Roediger, and McDermott (1999). There were five DRM lists of neutral words, ten negative lists, and ten positive lists, therefore participants viewed a total of 15 word lists. There were 12 words in each word list. Previous research investigating the effect of sleep on DRM false memories have used between 10 and 15 word lists, (eg., Diekelmann et al., 2008; Fenn et al., 2009), thus we chose 15 word lists to remain consistent with the literature, whilst also maximizing the potential between subjects effect of negative versus positive emotionality within the set of lists, with a 2:1 ratio of emotional versus neutral lists. A between subject design was used for positive versus negative word lists, to remove the possibility that viewing both negative and positive word lists together could reduce the overall valence of the word lists. Negative, positive and neutral word lists were selected such that word lists in each condition were matched as closely as possible on BAS, frequency, and arousal, and differed in ratings of valence. Despite this, negative and positive related lure words did differ in ratings of frequency (see Table 2).

Mean valence and arousal scores for word lists (see Table 1) and related lures (see Table 2) were taken from Warriner, Kuperman, and Brysbaert's (2013) affective norms of valence, arousal and dominance for 13,915 English words. Independent samples t-tests revealed that positive word lists significantly differed in valence from both neutral word lists $t(13) = .732$, SE = .202, $p = .001$, and negative lists $t(18) = 7.685$, SE = .312, $p < .001$. Negative word lists also significantly differed from neutral lists $t(13) = 3.617$, SE = .439, $p = .003$. Similarly, positive lure words significantly differed in valence from both neutral $t(13) = 3.765$, SE = .361, $p = .002$, and negative lures $t(18) = 10.056$, SE = .395, $p < .001$. Negative and neutral lures also significantly differed $t(13) = 4.462$, SE = .585, $p = .001$. For ratings of arousal, both positive and negative word lists significantly differed from neutral lists $t(13) = 2.235$, SE = .267, $p = .044$; $t(13) = 2.698$, SE = .252, $p = .018$. Positive and negative word lists did not significantly differ in arousal $t(18) = .378$, SE = .223, $p = .710$. For lure words, positive and negative lures did not significantly differ in arousal from neutral lure words $t(13) = 1.838$, SE = .872, $p = .089$; $t(13) = 1.189$, SE = .536, $p = .256$. Positive and negative lure words did not significantly differ in arousal $t(18) = .522$, SE = .450, p = .608.

Frequency scores for both word lists (Table 1) and lure words (Table 2) were taken from the van Heuven, Mandera, Keuleers, and Brysbaert's (2014) SUBTLEX-

UK word frequency norms for British English words, using the Zipf-values. Independent samples t-tests indicated that overall frequency of words did not significantly differ for positive and negative lists *t*(237) = 1.574, SE = .102, *p* = .117, positive and neutral lists *t*(177) = .902, SE = .123, *p* = .368, or negative and neutral lists *t*(178) = -.420, SE = .123, *p* = ..675. For lure words, positive and negative lure words did not significantly differ from neutral lures *t*(13) = .671, SE = .276, *p* = .514; *t*(13) = -1.346, SE = .266, *p* = .201. However, positive and negative lure words did differ in frequency *t*(18) = 2.248, SE = .242, *p* = .037, with positive lure words being rated as higher in frequency (see Table 2).

BAS ratings of word lists (Table 1) were taken from the University of South Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 2004). Independent samples t-tests indicated no significant difference in BAS between positive and negative word lists *t*(18) = -.490, SE = .056, *p* = .630, positive and neutral word lists *t*(13) = .319, SE = .064, *p* = .755, or negative and neutral word lists *t*(13) = .888, SE = .054, *p* = .391.

Table 4.1

Means and standard deviation of positive, negative and neutral word lists for valence, arousal, frequency and BAS.

	Positive	Negative	Neutral
Valence	6.643(.350)	4.242(.924)	5.830(.408)
Arousal	4.189(.516)	4.273(.480)	3.593(.414)
Frequency	4.484(.778)	4.323(.803)	4.375(.735)
BAS	0.240(.136)	0.267(.112)	0.219(.055)

Table 4.2

Means and standard deviation of positive, negative and neutral lure words for valence, arousal and frequency.

	Positive	Negative	Neutral
Valence	7.358(.519)	3.386(1.136)	5.998(.899)
Arousal	4.460(.929)	4.225(1.078)	3.588(.706)
Frequency	5.081(.552)	4.538(.529)	4.896(.371)

During testing 60 words were presented to participants, two previously seen words from each DRM word list (total of 30), the unseen, related lure word associated with each DRM list (total of 15), and 15 unseen, unrelated words. Each participant viewed the same two previously seen words, which were selected at random from the word lists. Unrelated words were taken either from lure words of unused DRM lists in Brainerd et al. (2010) or from Kousta, Vinson, and Vigliocco (2009), who developed emotional and neutral word lists derived from the ANEW database (Bradley & Lang, 1999). Unrelated words were matched to DRM word lists in terms of valence, resulting in five unrelated neutral words, ten unrelated negative words and ten unrelated positive words. All words were presented in Courier New bold, black font, lower case and in 18-point.

Participants in the sleep condition were required to wear an Actigraph sleep monitor overnight in order to measure their time spent asleep and the number of awakenings. Participants completed a questionnaire on sleep habits, caffeine and alcohol intake before each session, to ensure that they complied with instructions to sleep for at least 6 hours, to avoid excessive caffeine intake 12 hours before the first session and in between sessions, and to avoid alcohol for 12 hours before the first session and in between sessions.

2.1.1.3 Procedure

Participants were required to avoid alcohol and excessive caffeine consumption for 12 hours prior to the first session, as well as in-between sessions, which was assessed by a caffeine/alcohol intake questionnaire completed by participants at the start of each session. Participants were randomly allocated to one of four conditions, either wake negative, wake positive, sleep negative, or sleep positive, with 15 participants in each condition. Those in the wake conditions were exposed to DRM word lists at 9am and then tested the same day at 9pm. Those in the sleep conditions were trained on the word lists at 9pm, and were tested the following day at 9am. Those in the negative conditions were exposed to ten negative and five neutral DRM word lists, and those in the positive conditions were exposed to ten positive and the same five neutral DRM word lists.

During the first session, participants were exposed to the 15 lists of 12 words. Participants were seated approximately 60cm from the computer screen and were instructed to attempt to remember the words that appeared on the screen. Participants were first presented with a central fixation point for 500ms before words from one DRM list were presented one word at a time in the centre of the screen for 1500ms each. After each list, participants attempted to solve three simple maths problems that were presented for 10s each. The inclusion of maths problems acted as a distractor task, in order to prevent participants from rehearsing the word lists, as well as to ensure that lists were presented separately to provide context for the theme of each list. Maths problems were presented in a random order for each participant, and each problem was only presented once throughout the training session. After participants had viewed the three simple maths problems, the fixation cross re-appeared and participants were given another DRM word list to remember. The presentation order of word lists was randomised, and the order in which each word in each list was presented was also randomised.

Participants then left the lab, and participants in the wake condition were instructed not to nap throughout the day. Twelve hours later, after a period of daytime wakefulness or overnight sleep, participants returned for the testing phase of the study. During this second session, participants first completed another sleep habits and caffeine/alcohol questionnaire. Then, participants were presented with a central fixation cross for 500ms, before one of the test words was presented to participants in the centre of the screen. Words were presented for 120ms each, and the presentation order of each word was randomised for each participant. Participants were asked to decide whether they recognised the word as one that was presented during the training session or not. They did this through the press of a key on the keypad, with a press of zero corresponding to an old word (previously seen) and one corresponding to a new word (previously unseen). The numbers zero and one on the keypad were labelled ‘old’ and ‘new’ respectively. Participants were not given a response deadline. Once participants had made their decision, they were then presented with another fixation point for 500ms, before another word appeared on the screen.

2.1.1.4 Data Analysis

We measured both accuracy of responses, in terms of whether the word had appeared during training or not, and reaction time, though the latter can only be taken as suggestive as participants were not instructed to respond quickly to stimuli. If sleep affects memory for emotional versus neutral stimuli, then we expect to observe an interaction between emotion and sleep or wake group which may vary for veridical memory or for false memory generation. However, if sleep affects memory for positive versus negative emotional stimuli differently then we expect to observe an interaction between overall emotionality, valence, and sleep or wake groups which may be different for old, lure, and new words during testing. Thus, the critical statistical analysis to investigate is the four-way interaction in the results.

We analysed both response accuracy and response time (RT), using Linear Mixed-Effects Models (Baayen, 2008; Jaeger, 2008) in R (R Core Team, 2015). We took a hypothesis-driven approach to model building, thus, since we predicted a four-way interaction, we began with a maximal model including random effects of participants and items, as well as the four-way interaction between word type (old/lure/new), emotionality (emotional/neutral), valence (negative/positive), and group (sleep/wake). Exploratory analyses are reported separately.

2.1.2 Results and Discussion

2.1.2.1 Accuracy

We performed a Generalized Linear Mixed-Effects Model analysis (Baayen, 2008; Jaeger, 2008), modeling the probability (log odds) of response accuracy, including random effects of subjects and items (words), and a maximal model including fixed effects of group (sleep/wake), valence (negative/positive), emotionality (emotional/neutral) and word type (old/lure/new), and the interactions between these factors. Higher accuracy scores for old words indicates greater veridical memory performance, whereas higher accuracy for related lure words indicates fewer false memories. The full model is reported in Table 3, and mean accuracy for old, lure, and new words is show in Figure 1.

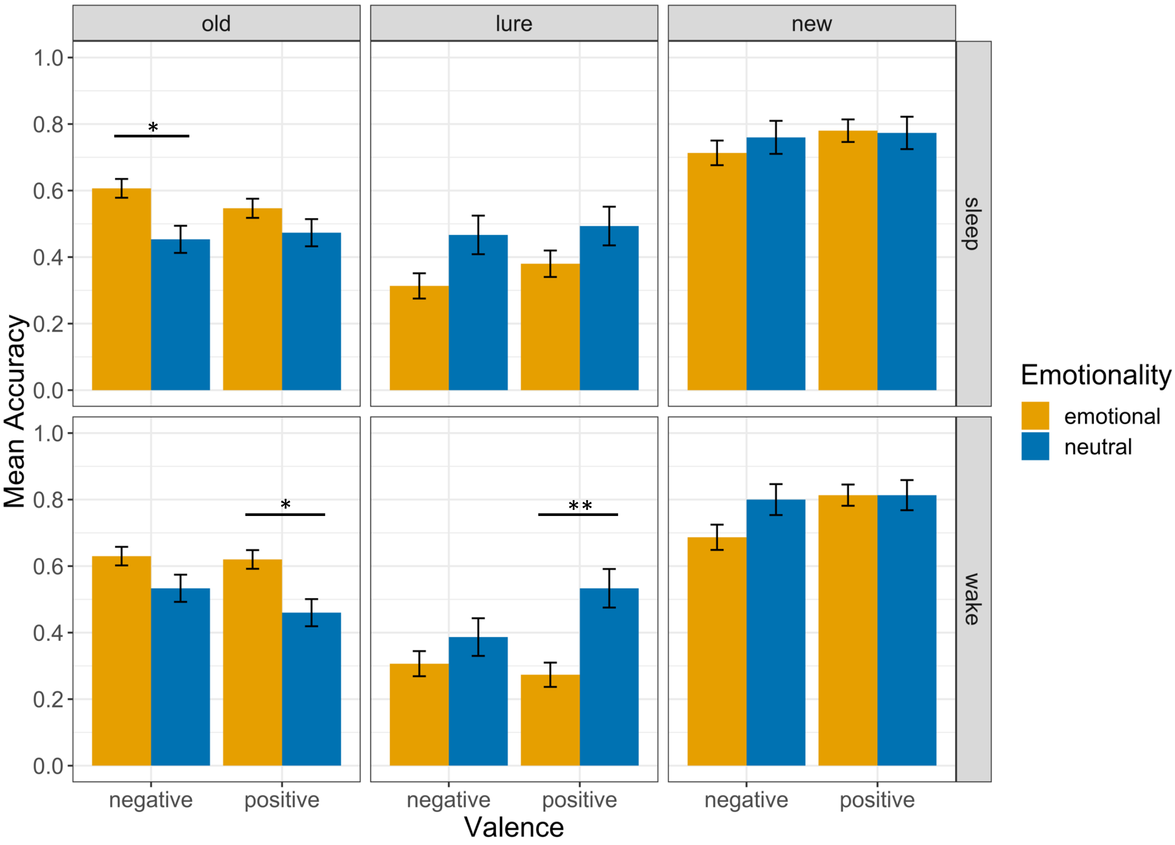


Figure 4.1. Mean accuracy performance for old, lure, and new words of negative, positive, and neutral valence in the sleep and wake groups, error bars show ±1 SEM, **p* < .05, ***p* < .01.

Table 4.3

Summary of general linear mixed effects model of response accuracy in experiment 1, including all fixed effects and all interaction terms.

	Estimated		Wald confidence intervals			
Fixed effects	coefficient	SE	2.50%	97.50%	z	Pr(>|z|)
(Intercept)	-0.822	0.228	-1.268	-0.376	-5.054	<.001
Word type (new vs. lure)	1.775	0.321	1.147	2.404	6.184	<.001
Word type (old vs. lure)	1.279	0.273	0.744	1.814	7.239	<.001
Emotionality (neutral vs. emotional)	0.685	0.376	-0.052	1.423	1.606	.069
Group (morning vs. evening)	-0.032	0.259	-0.540	0.476	1.059	.901
Valence (positive vs. negative)	0.312	0.317	-0.310	0.933	1.982	.325
Word type (new vs. lure):emotionality interaction	-0.410	0.555	-1.450	0.678	-0.085	.460
Word type (old vs. lure):emotionality interaction	-1.334	0.460	-2.235	-0.433	-2.462	.004
Word type (new vs. lure):group interaction	-0.101	0.362	-0.809	0.608	0.245	.781
Word type (old vs. lure):group interaction	0.137	0.308	-0.467	0.740	-0.645	.657
Emotionality:group interaction	-0.303	0.421	-1.128	0.522	-0.414	.471
Word type (new vs. lure):valence interaction	0.046	0.456	-0.848	0.940	0.241	.920
Word type (old vs. lure):valence interaction	-0.571	0.381	-1.318	0.175	-1.503	.134
Emotionality:valence interaction	-0.202	0.456	-1.095	0.690	-1.196	.657
Group:valence interaction	-0.476	0.366	-1.193	0.241	-2.107	.193
Word type (new vs. lure):emotionality:group interaction	0.683	0.638	-0.568	1.934	0.116	.285
Word type (old vs. lure):emotionality:group interaction	0.529	0.512	-0.474	1.533	1.321	.301
Word type (new vs. lure):emotionality:valence interaction	-0.076	0.689	-1.426	1.273	-0.449	.912
Word type (old vs. lure):emotionality:valence interaction	0.545	0.556	-0.546	1.635	1.015	.328
Word type (new vs. lure):group:valence interaction	0.820	0.528	-0.215	1.856	1.587	.121
Word type (old vs. lure):group:valence interaction	0.689	0.433	-0.161	1.539	1.210	.112
Emotionality:group:valence interaction	0.976	0.592	-0.185	2.137	1.042	.100
Word type (new vs. lure):emotionality:group:valence interaction	-1.309	0.915	-3.102	0.485	-1.241	.153
Word type (old vs. lure):emotionality:group:valence interaction	-1.574	0.721	-2.988	-0.160	-1.434	.029

Random effects	Name	Variance	Std. Dev
Subject effect on intercepts	(Intercept)	0.018	0.133
Item effect on intercepts	(Intercept)	0.182	0.426

	AIC	BIC	logLik	deviance
	4578.5	4739.4	-2263.2	4526.5

3660 observations, 60 participants, 60 items

Confirmatory Analyses

The results revealed a significant main effect of word type, indicating that participants exhibited false memory effects for the lure words, in line with previous studies using DRM lists. Old words were accurately accepted significantly more than unseen lure words were rejected, and new words were accurately rejected significantly more than lure words.

There was no significant main effect of emotionality, valence, or group. So overall memory accuracy for emotional words was similar to that of neutral words, overall memory for positive emotional words was similar to that of negative emotional words, and memory for the sleep and wake group participants was similar. These results suggest that the control of the stimuli with respect to the psycholinguistic variables was effective, and that the wake and sleep groups did not differ in their overall memory abilities.

The results did reveal a significant word type by emotionality interaction for old versus lure words, but not for lure versus new words or old versus new words. A further Generalized Linear Mixed-Effects Model only for old words with emotionality as a fixed effect, revealed greater accuracy for emotional than neutral words (estimate = -.538, SE = .201, z = -2.680, p = .007). An analysis only for the lure words resulted in the opposite effect, with emotional words less accurate than neutral words (estimate = .686, SE = .229, z = 3.002, p = .003), consistent with previous studies of negative emotion on false memory production (Howe et al., 2010). Further analyses on emotional and neutral words separately indicated that, for emotional words, there was a significant difference between old and lure words (estimate = 1.232, SE = .175, z = 7.054, $p = 1.74 \times 10^{-12}$), with old words being accurately accepted significantly more than lure words were accurately rejected. However, for neutral words, there was no significant difference in performance between old and lure words, (estimate = .040, SE = .156, z = .257, *p* = .797). No other two-way or three-way interactions were significant.

Critically, the results revealed a significant four-way interaction between word type, emotionality, valence, and group, which was driven by the contrast between old and lure words (see Figure 1). We initially ran a series of Generalized Linear Mixed-Effects Models for sleep group separately comparing old and lure words to determine how sleep differentially affected memory for positive and negative emotional words.

For the sleep group, we found the opposite effect of valence on memory performance to the wake group. A series of Generalized Linear Mixed-Effects Models indicated that for the positive group, there was no significant difference in performance accuracy between emotional and neutral words for the old words (estimate = -.332, SE = .319, z = -1.042, *p* = .298), or lure words (estimate = .463, SE = .286, z = 1.620, *p* = .105). Positive old words were responded to more accurately than positive lure words (estimate = .718, SE = .301, z = 2.384, *p* = .017), however neutral old and lure words did not significantly differ in response accuracy (estimate = -.080, SE = .283, z = -.283, *p* = .777). For those participants in the sleep group who were exposed to negative and neutral word lists, negative old words were remembered significantly better than neutral old words (estimate = -.686, SE = .317, z = -2.165, *p* = .030), with no significant difference between negative and neutral lure words (estimate = .685, SE = .378, z = 1.813, *p* = .070). Similar to the positive group, negative old words were remembered better than negative lure words (estimate = 1.371, SE = .362, z = 3.785, *p* < .001), and there was no significant difference in performance between neutral old and lure words, (estimate = -0.054, SE = .284, z = -.189, *p* = .850). Sleeping between exposure and testing resulted in better performance for recognising old negative emotion words compared to neutral words, but no difference between positive emotion and neutral words.

Exploratory Analyses

We had no prior predictions as to how daytime wake might differentially influence performance accuracy for emotional versus neutral old and lure words. Thus, we ran exploratory analyses in a similar way to the analysis of the sleep group. For the wake group, the model revealed a marginally significant 3-way interaction between emotionality, valence and word type (estimate = -1.036, SE = .544, z = -1.905, *p* = .057). Further analysis of the performance accuracy of participants in the positive and negative emotional stimuli groups separately revealed that positive old words were accurately responded to significantly more than neutral old words (estimate = -1.740, SE = .296, z = -2.497, *p* = .013), whereas neutral lure words were accurately responded to significantly more than positive lure words (estimate = 1.325, SE = .433, z = 3.063, *p* = .002). Furthermore, positive old words were accurately responded to significantly more than positive lure words, (estimate = 1.519, SE = .270, z = 5.618, $p = 1.94\text{x } 10^{-8}$), however there was no significant difference in performance between old and lure neutral words (estimate = -.313, SE = .404, z = -.774, *p* = .439). For those in the wake group who viewed negative and neutral word lists, there was no significant difference in performance between emotional and neutral words for either old (estimate = -.425, SE = .233, z = -1.824, *p* = .068) or lure words (estimate = -.370, SE = .302, z = 1.224, *p* = .221). However negative old words were accurately responded to significantly more than negative lure words (estimate = 1.378, SE = .249, z = 5.543, $p = 2.98 \text{ x } 10^{-8}$), and neutral old words were responded to significantly more accurately than neutral lure words, (estimate = .610, SE = .292, z = 2.087, *p* = .037). Thus, staying awake between exposure to the word lists and the recognition test resulted in a greater effect of false memories for the positive compared to the neutral words and better recognition of old positive words than old neutral words, but no difference for the negative words.

2.1.2.2 Response Time

Response times (RT) for correct responses were analysed to establish whether any speed-accuracy trade-off effects influenced the accuracy findings, as well as to test whether any of the conditions affected participants' speed of access to stored information. Following the same method as Monaghan et al., (2017), responses that were slower than 2500ms and faster than 150ms were removed from the analysis (96 responses), resulting in the omission of 4.59% of the data, and leaving 1930 responses for analysis.

There was non-normality in the RT distribution, thus based on the suggestions from Baayen, Feldman, and Schreuder (2006), we reduced skew by transforming latencies to $\log_{10}$(RT). We performed Linear Mixed-Effects Model analysis, predicting response time, with random effects of subjects and items (words), and fixed effects of group (sleep/wake), valence (negative/positive), emotionality (emotional/neutral) and word type (old/lure/new), and the interactions between these factors.

Based on suggestions from Baayen (2008), who indicate that a coefficient is a significant predictor if $|t| > 2$, we found no significant main effects or interaction effects (see Table 4 for the full model, and Figure 2 for means and standard error).

In summary, these results demonstrate that, overall, emotional words are remembered significantly better than neutral words, however emotional words are also more susceptible to false memories. The results also indicate that sleep has a different effect on memory of positive versus negative emotional words, boosting the consolidation of negative emotional words, leading to increased veridical memory in comparison to neutral words, but having no effect on memory of positive versus neutral words. In contrast, wake enhanced veridical and false memory for positive words in comparison to neutral words but had no different effect for negative emotional words compared to neutral words.

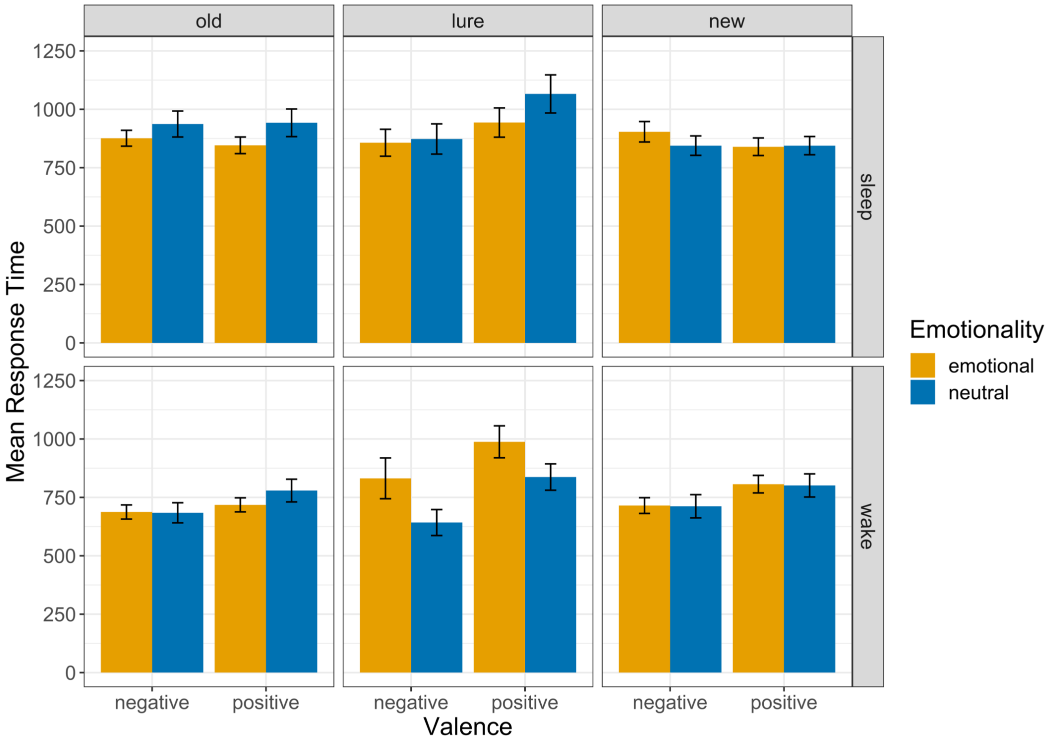


Figure 4.2. Mean response time to correct responses for old, lure, and new words of negative, positive, and neutral valence in the sleep and wake groups, error bars show ±1 SEM.

Table 4.4

Summary of linear mixed effects model of response time in experiment 1, including all fixed effects and all interaction terms.

Fixed effects	Estimated coefficient	SE	Wald confidence intervals 2.50%	97.50%	*t*
(Intercept)	2.906	0.039	2.829	2.982	74.371
Word type (new vs. lure)	-0.001	0.031	-0.062	0.061	-0.005
Word type (old vs. lure)	-0.004	0.029	-0.061	0.053	-0.142
Emotionality (neutral vs. emotional)	0.008	0.040	-0.070	0.086	0.199
Group (wake vs. sleep)	-0.103	0.055	0.212	0.005	-1.864
Valence (positive vs. negative)	0.035	0.054	-0.072	0.141	0.639
Word type (new vs. lure):emotionality interaction	-0.015	0.050	-0.112	0.083	-0.295
Word type (old vs. lure):emotionality interaction	0.013	0.047	-0.080	0.105	0.264
Word type (new vs. lure):group interaction	-0.026	0.043	-0.111	0.059	-0.592
Word type (old vs. lure):group interaction	-0.034	0.041	-0.114	0.046	-0.825
Emotionality:group interaction	-0.055	0.057	-0.166	0.056	-0.972
Word type (new vs. lure):valence interaction	-0.046	0.043	-0.129	0.038	-1.067
Word type (old vs. lure):valence interaction	-0.062	0.040	-0.141	0.017	-1.546
Emotionality:valence interaction	0.060	0.054	-0.045	0.165	1.117
Group:valence interaction	0.075	0.078	-0.077	0.227	0.964
Word type (new vs. lure):emotionality:group interaction	0.069	0.069	-0.065	0.204	1.012
Word type (old vs. lure):emotionality:group interaction	0.047	0.066	-0.082	0.175	0.712
Word type (new vs. lure):emotionality:valence interaction	-0.049	0.067	-0.179	0.082	-0.727
Word type (old vs. lure):emotionality:valence interaction	-0.052	0.064	-0.177	0.073	-0.816
Word type (new vs. lure):group:valence interaction	0.015	0.060	-0.103	0.132	0.244
Word type (old vs. lure):group:valence interaction	0.004	0.057	-0.108	0.116	0.076
Emotionality:group:valence interaction	-0.049	0.077	-0.199	0.102	-0.631
Word type (new vs. lure):emotionality:group:valence interaction	0.022	0.094	-0.162	0.207	0.238
Word type (old vs. lure):emotionality:group:valence interaction	0.052	0.091	-0.126	0.229	0.573

Random effects	Name	Variance	Std. Dev
Subject effect on intercepts	(Intercept)	0.013	0.113
Item effect on intercepts	(Intercept)	<.001	0.021

	AIC	BIC	logLik	deviance
	-1242.7	-1092.4	648.4	-1296.7

1930 observations, 60 participants, 60 items

2.2 Experiment 2: Effects of time of day on the processing of emotional and neutral word lists

As the first experiment tested the effects of overnight sleep in comparison to daytime wakefulness on memory performance, it could be the case that the difference in performance between the two groups may be due to time of day of training and testing. The wake group were trained in the morning and tested in the evening, whereas the sleep group were trained in the evening and tested in the morning. It is possible that there is a general increase in performance for tasks completed in the morning in comparison to the evening (Schmidt, Collette, Cajochen, & Peigneux, 2007), therefore differences between these two groups could be due to better task performance in terms of encoding during the morning for the wake group, or better recognition in the morning by the sleep group.

It is also possible that emotional states of participants varied from the morning to the evening sessions. If participants' mood was more negative in the morning than the evening then this could result in greater recognition memory effects for negative compared to neutral or to positive stimuli, due to mood congruency effects (Bland, Howe, & Knott, 2016; Howe & Malone, 2011; Ruci, Tomes, & Zelenski, 2009). In this control study, we therefore included a measure of mood to determine if that could explain different patterns of the emotional memory effects for the sleep and wake groups in Experiment 1.

Using a similar methodology to previous studies measuring the effects of sleep on the construction of false memories (Fenn et al., 2009; McKeon et al., 2012; Payne et al., 2008), we tested two control groups who were trained and tested in the same session at either 9am or 9pm, to investigate whether time of day had an impact on emotional memory processing.

2.2.1 Methods

2.2.1.1 Participants

Sixty-one Lancaster University students (47 female, 14 male) with a mean age of 20.25 years (SD = 3.05, range = 18 to 31) took part in the study for course credit or for payment of £6.50. All participants gave informed consent and were fully debriefed at the end of the study. An additional 12 participants were tested but excluded, due to being non-native English speakers (*n* = 7), aged over 31 years old (*n* = 2), or not sleeping the night before the experiment took place (*n* = 3).

2.2.1.2 Materials

The same stimuli used in Experiment 1 were employed in this study. In addition, participants were given the Self-Assessment Manikin scale (SAM; Bradley & Lang, 1994) in order to assess their mood at that time. The SAM is a pictorial assessment measure of mood, whereby participants rate their feelings of pleasure, arousal, and dominance on a scale from 1-9.

2.2.1.3 Procedure

Participants were randomly allocated to one of four conditions, either morning exposed to the negative and neutral word lists, morning and exposed to positive and neutral word lists, evening and negative and neutral lists, or evening and positive and neutral lists. Exposure to word lists and subsequent recognition testing occurred in a single session. Those in the morning conditions took part in the study at 9am, and those in the evening conditions took part at 9pm. First, participants completed a questionnaire on their sleep habits, caffeine and alcohol intake, and were given the SAM mood questionnaire. They were then presented with 15 lists of 12 words as in Experiment 1, following the same procedure as Experiment 1. Participants were then asked to complete sudokus during a 20-minute break. The delay duration was based on previous studies that have used a similar time period between sessions in their control conditions (McKeon et al., 2012; Payne et al., 2009). Participants then answered a second SAM mood questionnaire, and took part in the recognition task, again following the same procedure as in Experiment 1.

2.2.1.4 Data Analysis

In Experiment 2, we wanted to directly compare the findings from Experiment 1, therefore for both accuracy and RT we again ran linear mixed effects models, starting with the same maximal model structure as in Experiment 1, and removing interactions and main effects if they did not significantly improve model fit. We hypothesized that we would find no difference between groups (morning/evening), and thus predicted no four-way interaction. As in Experiment 1, we predicted a main effect of word type, and an interaction between word type (old versus lure) and emotionality. The analyses are split accordingly to reflect hypothesis-driven versus further exploratory analyses.

2.2.2 Results

2.2.2.1 Accuracy

We followed the same analysis as in Experiment 1, testing a series of Generalized Linear Mixed-Effects Models, modeling the probability (log odds) of response accuracy considering both the random variation across participants and items (words), as well as the fixed effects of group (morning/evening), valence (negative/positive), emotionality (emotional/neutral) and word type (old/lure/new), and the interactions between these effects. The full model is reported in Table 5, and descriptive statistics are presented in Figure 3.

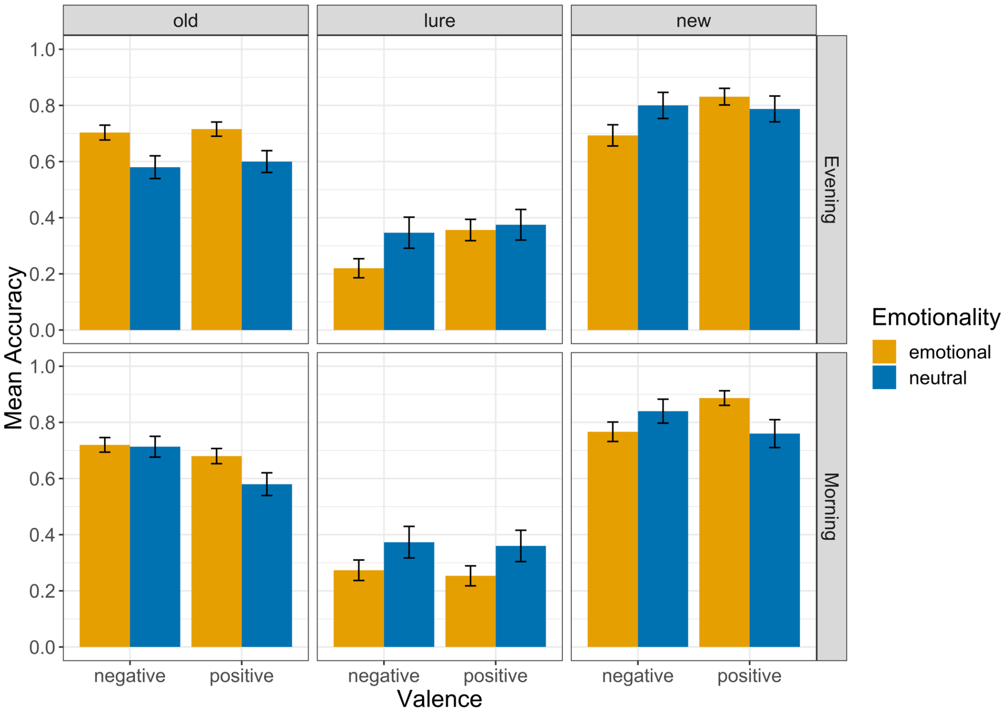


Figure 4.3. Mean performance accuracy for old, lure, and new words of negative, positive, and neutral valence in the morning and evening groups, error bars show ±1 SEM.

Table 4.5

Summary of general linear mixed effects model of response accuracy in experiment 2, including all fixed effects and all interaction terms.

	Estimated		Wald confidence intervals			
Fixed effects	coefficient	SE	2.50%	97.50%	z	Pr(>|z|)
(Intercept)	-1.339	0.265	-1.859	-0.820	-5.054	<.001
Word type (new vs. lure)	2.207	0.357	1.508	2.907	6.184	<.001
Word type (old vs. lure)	2.268	0.313	1.654	2.882	7.239	<.001
Emotionality (neutral vs. emotional)	0.683	0.425	-0.151	1.517	1.606	.108
Group (morning vs. evening)	0.301	0.285	-0.256	0.859	1.059	.290
Valence (positive vs. negative)	0.711	0.359	0.008	1.414	1.982	.047
Word type (new vs. lure):emotionality interaction	-0.053	0.619	-1.266	1.161	-0.085	.932
Word type (old vs. lure):emotionality interaction	-1.272	0.517	-2.284	-0.259	-2.462	.014
Word type (new vs. lure):group interaction	0.094	0.383	-0.658	0.845	0.245	.807
Word type (old vs. lure):group interaction	-0.214	0.331	-0.863	0.436	-0.645	.519
Emotionality:group interaction	-0.183	0.441	-1.047	0.682	-0.414	.679
Word type (new vs. lure):valence interaction	0.122	0.506	-0.870	1.114	0.241	.810
Word type (old vs. lure):valence interaction	-0.644	0.428	-1.483	0.195	-1.503	.133
Emotionality:valence interaction	-0.584	0.488	-1.541	0.373	-1.196	.232
Group:valence interaction	-0.824	0.391	-1.590	-0.057	-2.107	.035
Word type (new vs. lure):emotionality:group interaction	0.079	0.677	-1.247	1.405	0.116	.907
Word type (old vs. lure):emotionality:group interaction	0.713	0.540	-0.345	1.771	1.321	.187
Word type (new vs. lure):emotionality:valence interaction	-0.329	0.734	-1.768	1.109	-0.449	.654
Word type (old vs. lure):emotionality:valence interaction	0.605	0.596	-0.563	1.772	1.015	.310
Word type (new vs. lure):group:valence interaction	0.906	0.571	-0.213	2.025	1.587	.113
Word type (old vs. lure):group:valence interaction	0.553	0.457	-0.343	1.449	1.210	.226
Emotionality:group:valence interaction	0.639	0.613	-0.562	1.839	1.042	.297
Word type (new vs. lure):emotionality:group:valence interaction	-1.183	0.953	-3.051	0.685	-1.241	.214
Word type (old vs. lure):emotionality:group:valence interaction	-1.074	0.749	-2.542	0.394	-1.434	.152

Random effects	Name	Variance	Std. Dev
Subject effect on intercepts	(Intercept)	0.042	0.206
Item effect on intercepts	(Intercept)	0.266	0.516

	AIC	BIC	logLik	deviance
	4276.4	4437.7	-2112.2	4224.4

3660 observations, 61 participants, 60 items

Confirmatory Analyses

The results demonstrated a significant main effect of word type, similar to Experiment 1. Old words were accurately accepted significantly more than lure words were rejected, and new words were accurately rejected significantly more than lure words. The results also revealed a significant main effect of valence, with those viewing positive and neutral word lists performing better than those who viewed negative and neutral DRM lists. We found no significant main effect of emotionality or group.

As with Experiment 1, the results also indicated a significant word type by emotionality interaction for old versus lure words. Similar to Experiment 1, we therefore analysed old and lure words separately. For old words the difference between emotional and neutral word lists was close to significance, (estimate = -.441, SE = .225, z = -1.958, *p* = .050); emotional words were responded to more accurately than neutral words. For lure words, performance between emotional and neutral words was not significantly different (estimate = .466, SE = .274, z = 1.704, *p* = .088). Further analyses on emotional and neutral words separately indicated that, for emotional words, there was a significant difference between old and lure words (estimate =1.969, SE = .196, z = 10.031, $p = 2 \times 10^{-16}$); with old words being accurately accepted significantly more than lure words were accurately rejected. Neutral old words were also responded to significantly more accurately than neutral lure words (estimate = .1.065, SE = .213, z = .4.995, $p = 5.88 \times 10^{-7}$).

Exploratory Analyses

Distinct from Experiment 1, we also found a significant group by valence interaction (see Figure 4). Analysis of the morning and evening groups separately revealed that, for those in the evening group, there was no significant difference in accuracy between participants who viewed negative and neutral or positive and neutral word lists (estimate = .166, SE = .166, z = .995, *p* = .320). Similarly, there was no significant difference between negative and positive groups in the morning group, (estimate = -.310, SE = .173, z = -1.799, *p* = .072). Investigating the positive and negative groups separately revealed no significant difference between morning and evening groups for those exposed to positive and neutral word lists, (estimate = -.142, SE = .126, z = -1.131, *p* = .258). However, for those who viewed negative and neutral word lists, the morning group responded with significantly greater accuracy than the evening group (estimate = .285, SE = .135, z = 2.116, *p* = .034).

Therefore, those participants who encode and retrieve negative and neutral word lists in the evening showed a reduction in performance accuracy compared to those in the morning condition. The results of Experiment 1 reveal a bias for negative emotion after sleep, and a bias for positive emotion after wake. Thus, the results of Experiment 2 are not sufficient to explain the differences between the sleep and wake group performance that was evident in Experiment 1, since there was no bias in encoding and retrieving negative emotional words when negative words were encoded in the morning in the wake group or retrieved in the morning in the sleep group.

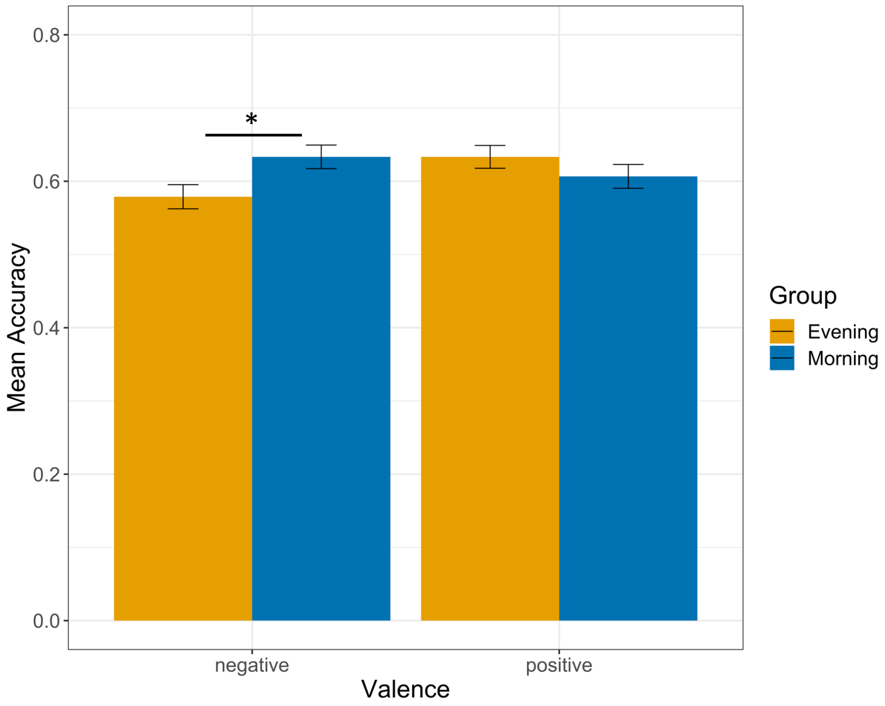


Figure 4.4. Mean accuracy performance in the morning compared to the evening for those participants who viewed negative and neutral lists, and those who viewed positive and neutral lists, error bars show ±1 SEM, **p* < .05.

2.2.2.2 Response Time

As in experiment 1, RTs for correct responses were analysed. Again, responses that were slower than 2500ms and faster than 150ms were removed from the analysis (154 responses), resulting in the omission of 6.86% of the data, and leaving 2091 responses for analysis. For those in the evening groups, we removed 56 responses from the negative condition, and 28 responses from the positive condition. For those in the morning groups, we removed 37 responses from the negative condition, and 33 responses from the positive condition. As with experiment 1, we reduced skew in the RT distribution by transforming latencies to $\log_{10}$(RT). We again performed a Linear Mixed-Effects Model analysis, modeling response time, with random effects of subjects and items (words), and fixed effects of group (morning/evening), valence (negative/positive), emotionality (emotional/neutral) and word type (old/lure/new), and the interactions between these factors.

The analysis revealed a main effect of word type; old words were accurately accepted significantly quicker than lure words were rejected. No other main effects or interaction effects were found (see Table 6 for the full model including inferential statistics, see Figure 5 for means and standard error).

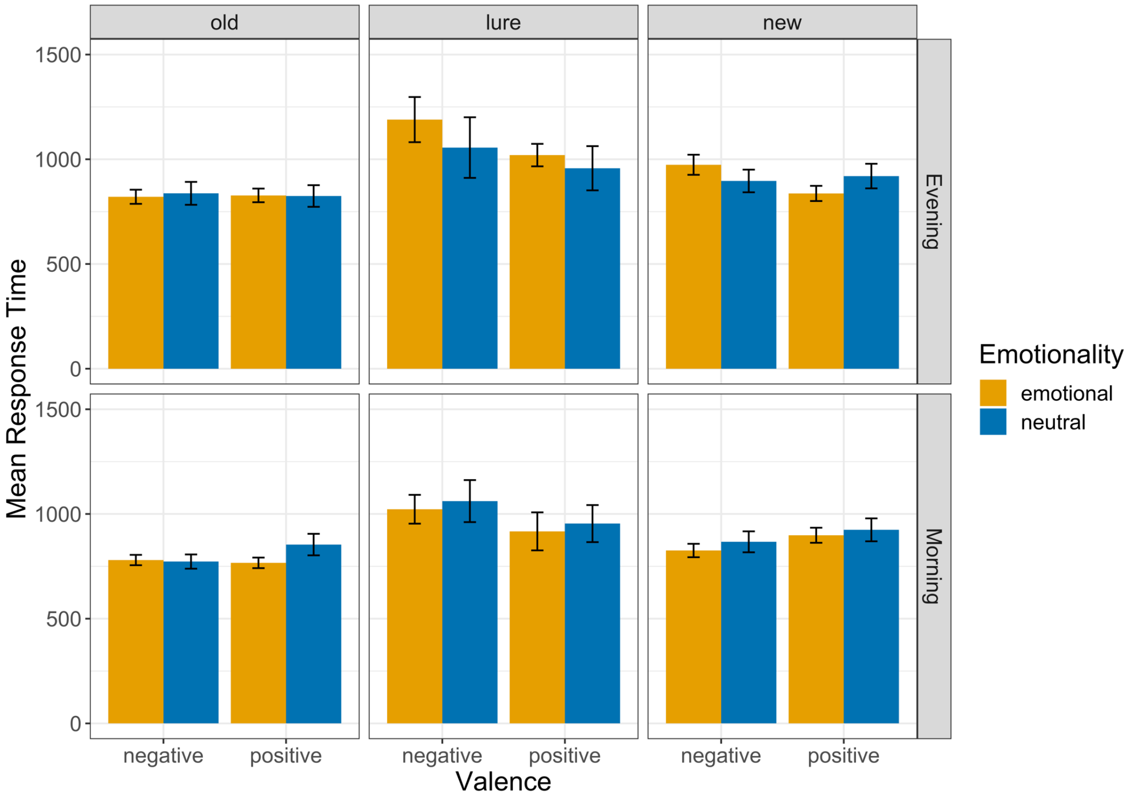


Figure 4.5. Mean response time to correct responses for old, lure, and new words of negative, positive, and neutral valence in the sleep and wake groups, error bars show ±1 SEM.

Table 4.6

Summary of linear mixed effects model of response time in experiment 2, including all fixed effects and all interaction terms.

	Estimated		Wald confidence intervals		
Fixed effects	coefficient	SE	2.50%	97.50%	*t*
(Intercept)	2.997	0.044	2.912	3.083	68.777
Word type (new vs. lure)	-0.056	0.035	-0.125	0.013	-1.602
Word type (old vs. lure)	-0.139	0.033	-0.204	-0.074	-4.212
Emotionality (neutral vs. emotional)	-0.054	0.050	-0.151	0.043	-1.084
Group (wake vs. sleep)	-0.012	0.059	-0.128	0.105	-0.198
Valence (positive vs. negative)	-0.050	0.058	-0.163	0.063	-0.865
Word type (new vs. lure):emotionality interaction	0.028	0.058	-0.085	0.141	0.494
Word type (old vs. lure):emotionality interaction	0.062	0.055	-0.046	0.169	1.121
Word type (new vs. lure):group interaction	-0.058	0.046	-0.148	0.032	-1.268
Word type (old vs. lure):group interaction	-0.003	0.043	-0.087	0.082	-0.060
Emotionality:group interaction	0.051	0.064	-0.074	0.176	0.796
Word type (new vs. lure):valence interaction	-0.004	0.045	-0.092	0.084	-0.088
Word type (old vs. lure):valence interaction	0.054	0.042	-0.028	0.136	1.285
Emotionality:valence interaction	0.034	0.062	-0.088	0.155	0.547
Group:valence interaction	-0.020	0.082	-0.181	0.140	-0.249
Word type (new vs. lure):emotionality:group interaction	-0.002	0.074	-0.147	0.143	-0.032
Word type (old vs. lure):emotionality:group interaction	-0.060	0.070	-0.198	0.078	-0.853
Word type (new vs. lure):emotionality:valence interaction	0.022	0.073	-0.120	0.165	0.308
Word type (old vs. lure):emotionality:valence interaction	-0.035	0.069	-0.170	0.100	-0.505
Word type (new vs. lure):group:valence interaction	0.104	0.062	-0.017	0.225	1.686
Word type (old vs. lure):group:valence interaction	0.010	0.058	-0.105	0.124	0.167
Emotionality:group:valence interaction	0.008	0.086	-0.161	0.176	0.087
Word type (new vs. lure):emotionality:group:valence interaction	-0.076	0.100	-0.272	0.121	-0.754
Word type (old vs. lure):emotionality:group:valence interaction	0.036	0.095	-0.151	0.223	0.375

Random effects	Name	Variance	Std. Dev
Subject effect on intercepts	(Intercept)	0.015	0.121
Item effect on intercepts	(Intercept)	<.001	0.02

	AIC	BIC	logLik	deviance
	-1487.7	-1332.3	769.4	-1538.7

2091 observations, 61 participants, 60 items

2.2.2.3 Self-Assessment Manikin

Participants filled out a SAM mood questionnaire twice during the study, the first before taking part in the training phase, the second before taking part in the recognition test. The SAM mood questionnaire assessed participant's feelings of valence (ranging from negative-positive), arousal (from calm-excited), and control (dominated-dominant). Mean scores of valence, arousal, and control are reported in Table 7, and show no difference in ratings of mood between participants exposed to positive and neutral, or negative and neutral word lists, or between those in the morning and evening group.

Table 4.7

Means and standard deviations of ratings of valence, arousal, and control split between morning and evening groups, and for participants exposed to negative and neutral, and positive and neutral word lists.

	Group		Valence	
	Morning	Evening	Negative	Positive
Mean (SD) valence mood score	6.533 (0.900)	6.274 (1.132)	6.600 (1.155)	6.210 (0.854)
Mean (SD) arousal mood score	4.300 (1.284)	4.000 (1.329)	4.217 (1.442)	4.081 (1.177)
Mean (SD) control mood score	4.917 (1.267)	5.129 (1.176)	4.967 (1.181)	5.081 (1.266)

Despite this, to ensure that mood had no effect on overall performance accuracy, we performed Generalized Linear Mixed-Effects Models with recognition memory accuracy as the dependent variable, which demonstrated that adding SAM valence score did not improve model fit compared to a model with only random effects, ($\chi^2(1) = .059$, $p = .808$), nor did adding SAM arousal score, ($\chi^2(1) = .753$, $p = .386$), or SAM control score, ($\chi^2(1) = .021$, $p = .885$). This suggested no overall significant difference in accuracy dependent on ratings of mood.

Adding an interaction between group (morning/evening) and either valence score, ($\chi^2(3) = 2.857, p = .414$), arousal score, ($\chi^2(3) = 2.883, p = .410$), or control score, ($\chi^2(3) = 2.430, p = .488$), did not improve model fit compared to a model with only random effects. Similarly, adding the interaction term valence (positive/neutral or negative/neutral list exposure group) by valence score, ($\chi^2(3) = 5.177, p = .159$), arousal score, ($\chi^2(3) = 3.241, p = .356$), or control score, ($\chi^2(3) = 1.299, p = .729$) did not improve model fit. This indicates that participants' mood in the morning and evening groups and negative and positive valence groups did not significantly influence their overall memory performance accuracy.

These results suggest that the different effects of sleep and wake on subsequent veridical and false recognition for word lists of negative and positive valence are not due to differences in time-of-day of encoding or retrieval of these memories. However, the results of Experiment 2 show that participants perform significantly better on negative and neutral word lists in the morning than the evening. Since there was no difference in mood between the two groups, these results demonstrate the slight bias for encoding and retrieving negative emotional words in the morning compared to the evening found in Experiment 2 was not due to mood-congruency. However, importantly the two-way interaction between valence and group found in Experiment 2 was not evident in Experiment 1, meaning that the bias for encoding and retrieving negative emotional words was not evident when negative words were only encoded (as in the wake group) or retrieved (as in the sleep group) in the morning session.

3. General Discussion

A main aim of this study was to investigate the effects of emotionality of word lists on subsequent memory performance, and to determine to what extent both negative *and* positive emotionality influences the formation of memories for words in comparison to neutral, non-emotional word lists. Based on previous research, we aimed to investigate two specific hypotheses relating to emotion and memory. First, that emotionality of a word should increase accurate memory performance (Adelman & Estes, 2013). This was supported by the current findings, which indicate greater performance accuracy of previously seen emotional words compared to neutral words in both Experiment 1 and 2. A second hypothesis predicted that emotionality of the word lists would affect false recognition of lure words. Based on previous research (El Sharkawy et al., 2008; Howe et al., 2010), we predicted that the emotionality of word lists would elicit increased false recognition of emotional lure words compared to neutral words. This was also supported by the current study, which indicated an increase in false recognition of emotional lure words of both positive and negative valence, compared to neutral, non-emotional word lists. However, there was no significant difference found between false recognition of negative and positive lure words, which did not replicate the results of the research conducted by Brainerd et al. (2008). They suggested that, compared to words of neutral valence, negative valence leads to increasing associations of critical lures to the initial lists, increasing false recognition, whereas positive valence reduces the false memory effect. The current study, rather, indicated that there may be an overall influence of emotionality on increasing associations, leading to a general increase in false memories for emotional words of both negative and positive valence compared to neutral words.

A second main aim of the current study was to investigate the effects of sleep on emotional and non-emotional memory processing. Based on previous research, we aimed to investigate three specific hypotheses relating to sleep and emotional memory performance. First, previous studies indicate that sleep plays a different role in memory processing on tests of false recognition in comparison to false recall (Payne et al., 2009; Fenn et al., 2009). Based on this previous research, we predicted that there would be no overall difference in false recognition of lure words for the sleep and wake group. This was supported by the current findings, which indicated no overall effect of sleep on false memory performance, suggesting that recognition testing is much less vulnerable to post-sleep false memories, potentially due to increased source monitoring abilities (Fenn et al., 2009; Johnson, Hashtroudi, & Lindsay, 1993).

Second, it was hypothesised that sleep would both improve accurate recognition of previously seen words of negative valence, and increase false recognition of negative lure words, in comparison to neutral words. This could be due to the positive effect that REM sleep has on emotional memory consolidation (Carr & Nielsen, 2015; Nishida et al., 2009), and would support an adaptive memory system, that particularly protects and stores emotional memories (van der Helm & Walker, 2009; Walker & Stickgold, 2004). This was partially supported by the current research, which found greater accuracy for negative than neutral previously seen words after sleep, however no significant difference in performance between negative and neutral lure words was found. Since tests of false recognition after sleep show a reduction in lure word acceptance (Fenn et al., 2009), and tests of false recognition of emotional content shows an increase in lure word acceptance (El Sharkawy et al., 2008; Howe et al., 2010), it could be that the observed effect of negative versus neutral word lists on false recognition is weakened by a reduction in false recognition after sleep, thus explaining why we found no difference in performance between negative and neutral lure words after sleep. Furthermore, the findings of this study do not support the results of the only other study to date that investigates the effect of sleep on the processing of (negative) emotional and neutral false memories (McKeon et al., 2012). McKeon et al. indicated an increase in false recall of neutral lure words in comparison to negative lure words after sleep, supporting previous research to suggest a reduction in false recall of negative emotional lure words (Howe et al., 2010). This difference between the two studies further supports the possible differences between tests of recall and recognition of emotional and neutral word lists found (Howe et al., 2010), as well as the differences in false recall and recognition after sleep (Newbury & Monaghan, 2018). However, a potential limitation of these findings is that in order to examine the exact mechanisms occurring during sleep for emotional memory, the four-way interaction found led to multiple significance tests of the sleep and wake groups separately, and emotional and non-emotional words separately. It is therefore possible that some of these effects may not be replicated in future work with greater power. Further research should thus aim to replicate these findings with a greater sample size, to give increased power to the findings.

Finally, we did not know in advance precisely what effect sleep would have on the processing of words of positive valence, however based on previous research conducted by Brainerd et al., (2008), who found an increase in false recognition of negative lures, but a decrease in positive lure word false recognition, we suggested that we might also find a decrease in false recognition of positive lure words. This was not supported by the current findings. Instead, exploratory analyses found that word lists of positive valence led to a significant increase in accuracy and false recognition compared to neutral word lists, and this effect was only significant in the wake group. There was no significant difference in performance between positive and neutral word lists in the sleep group. Since these findings were exploratory, further research is needed to examine whether the effect of wake on emotional memory can be replicated both in DRM tasks as well as more broadly for memory performance. Furthermore, potential limitation of the current study is that we did not collect ratings of mood for Experiment 1, thus this difference between wake and sleep groups in emotional memory performance could be due to differences in mood between groups. However, since we did not find any differences in mood between morning and evening groups in Experiment 2, nor effects of mood on memory recognition performance, it is unlikely that such difference would account for the results of Experiment 1.

Another possible explanation for this increase in both accurate and false recognition of positive over neutral words after wake could be the fading affect bias (FAB; Walker, Skowronski & Thompson, 2003). This theory is based on research that indicates that in general, people are more likely to remember positive events in comparison to negative events over time (Ritchie et al., 2015; Walker, Vogl, & Thompson, 1997). Alternatively, Taylor (1991) suggested that we are less likely to remember negative events due to the mobilization-minimization hypothesis. This model suggests that after a negative event, we have heightened responses at the physiological, emotional and cognitive level. These responses are termed ‘mobilization’, and are then followed by cognitive and behavioural responses that aim to reduce the influence or effect of the event, termed ‘minimization’. This minimization of negative over neutral and positive events leads to reduced memory for negative events after weeks or months. Although this theory is based on memories for autobiographical events as opposed to memory for emotional words, it could be possible that a similar effect is occurring in the current study. Gibbons, Lee, and Walker (2011) found that the FAB effect occurs even within 12 hours of the event, and so this model could relate to our data by indicating that during wakefulness, participants work to minimize the negative emotional words to a greater extent than the positive emotional words. Since we did not see this bias for positive emotion during sleep, it could be suggested that the bias for positive emotional information during wake due to the FAB effect shifts to a bias for negative emotional information during sleep due to the offline processes involved in emotional memory consolidation (Carr & Nielsen, 2015; Nishida et al., 2009).

The findings of this study add to a growing field of research to show that emotionality not only improves veridical memory, but also leads to an increase in false recognition of emotional lure words. Furthermore, this study indicates that sleep plays an enhanced role in the processing of emotionally negative word lists, with greater veridical recognition of negative emotional words compared to neutral words after sleep, whereas during wake there is greater veridical and false recognition of positive emotional words compared to neutral words. Future research should focus on the extent to which these findings can be generalised to different types of emotional information, such as face recognition and memory for emotional events. Similarly, research investigating how emotional information is processed is particularly important to aid in minimizing the effects of negative emotional events, such as in those with post-traumatic stress disorder (PTSD) or depression. Increased false memories have been shown both in those suffering with depression (Jelinek, Hottenrott, Randjbar, Peters, & Moritz, 2009) and those with PTSD when they are exposed to trauma related information (Brennen, Dybdahl, & Kapidžić, 2007; Zoellner, Foa, Brigidi, & Przeworski, 2000). This will help us to understand the way in which different types of emotional and non-emotional information is encoded and processed, and the role that sleep plays in the consolidation and subsequent retrieval of these emotional memories.

References

Adelman, J. S., & Estes, Z. (2013). Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition, 129(3),* 530-535. doi:10.1016/j.cognition.2013.08.014

Baayen, R. H. (2008). *Analyzing linguistic data [electronic resource] : a practical introduction to statistics using R*. Cambridge: Cambridge : Cambridge University Press.

Bauer, L. M., Olheiser, E. L., Altarriba, J., & Landi, N. (2009). Word type effects in false recall: Concrete, abstract, and emotion word critical lures. *The American Journal of Psychology, 122*(4), 469-481.

Bland, C. E., Howe, M. L., & Knott, L. (2016). Discrete emotion-congruent false memories in the DRM paradigm. *Emotion, 16*(5), 611-619. doi:10.1037/emo0000153

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49-59. doi:10.1016/0005-7916(94)90063-9

Brainerd, C. J., Holliday, R., Reyna, V., Yang, Y., & Toglia, M. (2010). Developmental reversals in false memory: Effects of emotional valence and arousal. *Journal of Experimental Child Psychology, 107*(2), 137-154. doi:10.1016/j.jecp.2010.04.013

Brainerd, C. J., Stein, L., Silveira, R., Rohenkohl, G., & Reyna, V. (2008). How does negative emotion cause false memories? *Psychological Science, 19*(9), 919-925. doi:10.1111/j.1467-9280.2008.02177.x

Brennen, T., Dybdahl, R., & Kapidžić, A. (2007). Trauma-related and neutral false memories in war-induced Posttraumatic Stress Disorder. *Consciousness and Cognition, 16*(4), 877-885. doi:10.1016/j.concog.2006.06.012

Cai, D., Mednick, S. A., Harrison, E, M., Kanady, J. C., & Mednick, S. C. (2009). REM, not incubation, improves creativity by priming associative networks. *Proceedings of the National Academy of Sciences, 106*, 10130–10134. doi:10.1073/pnas.0900271106

Carr, M., & Nielsen, T. (2015). Morning rapid eye movement sleep naps facilitate broad access to emotional semantic networks. *Sleep, 38*(3), 433-443. doi:10.5665/sleep.4504

Chambers, A., & Payne, J. D. (2014). Laugh yourself to sleep: memory consolidation for humorous information. *Experimental Brain Research, 232*(5), 1415-1427. doi:10.1007/s00221-013-3779-7

Dennis, N. A., Kim, H., & Cabeza, R. (2007). Effects of aging on true and false memory formation: An fMRI Study. *Neuropsychologia, 45*(14), 3157-3166. doi:10.1016/j.neuropsychologia.2007.07.003

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience, 11*(2), 114-126. doi:10.1038/nrn2762

Diekelmann, S., Born, J., & Wagner, U. (2010). Sleep enhances false memories depending on general memory performance. *Behavioural Brain Research, 208*(2), 425-429. doi:10.1016/j.bbr.2009.12.021

Diekelmann, S., Landolt, H., Lahl, O., Born, J., & Wagner, U. (2008). Sleep loss produces false memories. *PloS one, 3*(10), e3512-e3512. doi:10.1371/journal.pone.0003512

El Sharkawy, J., Groth, K., Vetter, C., Beraldi, A., & Fast, K. (2008). False memories of emotional and neutral words. *Behavioural Neurology, 19*(1/2), 7-11. doi: 10.1155/2008/587239

Fenn, K. M., Gallo, D. A., Margoliash, D., Roediger, H. L., & Nusbaum, H. C. (2009). Reduced false memory after sleep. *Learning & Memory, 16*(9), 509-513. doi: 10.1101/lm.1500808

Gallo, D. A., & Roediger, H. L. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language, 47*(3), 469-497. doi:10.1016/S0749-596X(02)00013-X

Gibbons, J. A., Lee, S. A., & Walker, W. R. (2011). The fading affect bias begins within 12 hours and persists for 3 months. *Applied Cognitive Psychology, 25*(4), 663-672. doi:10.1002/acp.1738

Goldstein, A. N., & Walker, M. P. (2014). The role of sleep in emotional brain function. *Annual Review of Clinical Psychology, 10*, 679-708. doi:10.1146/annurev-clinpsy-032813-153716

Howe, M. L., Candel, I., Otgaar, H., Malone, C., & Wimmer, M. C. (2010). Valence and the development of immediate and long-term false memory illusions. *Memory, 18*(1), 58-75. doi:10.1080/09658210903476514

Howe, M. L., & Malone, C. (2011). Mood-congruent true and false memory: Effects of depression. *Memory, 19*(2), 192-201. doi:10.1080/09658211.2010.544073

Howe, M. L., Wimmer, M. C., Gagnon, N., & Plumpton, S. (2009). An associative-activation theory of children's and adults' memory illusions. *Journal of Memory and Language, 60*(2), 229-251. doi: 10.1016/j.jml.2008.10.002

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

Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (Transformation or Not) and towards Logit Mixed Models. *Journal of Memory and Language, 59*(4), 434-446. doi:10.1016/j.jml.2007.11.007

Jelinek, L., Hottenrott, B., Randjbar, S., Peters, M. J., & Moritz, S. (2009). Visual false memories in post-traumatic stress disorder (PTSD). *Journal of Behavior Therapy and Experimental Psychiatry, 40*(2), 374-383. doi:10.1016/j.jbtep.2009.02.003

Johnson, M., Hashtroudi, S., Lindsay, D., & Steinberg, Robert J. (1993). Source Monitoring. *Psychological Bulletin, 114*(1), 3-28. doi: 10.1037/0033-2909.114.1.3

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition, 31*(8), 1169-1180. doi:10.3758/BF03195800

Kensinger, E. A., & Corkin, S. (2004a). Two routes to emotional memory: distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences of the United States of America, 101*(9), 3310-3315. doi:10.1073/pnas.0306408101

Kensinger, E. A., & Corkin, S. (2004b). The effects of emotional content and aging on false memories. *Cognitive, Affective, & Behavioral Neuroscience, 4*(1), 1-9. doi:10.3758/CABN.4.1.1

Kousta, S. T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition, 112*(3), 473-481. doi:10.1016/j.cognition.2009.06.007

Marshall, L., & Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends in Cognitive Sciences, 11*(10), 442-450. doi:10.1016/j.tics.2007.09.001

McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language, 35*(2), 212-230. doi:10.1006/jmla.1996.0012

McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience, 27*(1), 1-28. doi:10.1146/annurev.neuro.27.070203.144157

McKeon, S., Pace-Schott, E. F., & Spencer, R. M. (2012). Interaction of sleep and emotional content on the production of false memories. *PloS one, 7*(11), e49353. doi:10.1371/journal.pone.0049353

Monaghan, P., Shaw, J. J., Ashworth-Lord, A., & Newbury, C. R. (2017). Hemispheric processing of memory is affected by sleep. *Brain and Language, 167*, 36-43. doi:10.1016/j.bandl.2016.05.003

Newbury, C. R., & Monaghan, P. (2018). When does sleep affect veridical and false memory consolidation?: A meta-analysis. *Psychonomic Bulletin & Review*, 1-14.

Nelson, D., McEvoy, C., & Schreiber, T. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 402-407. doi:10.3758/BF03195588

Nishida, M., Pearsall, J., Buckner, R. L., & Walker, M. P. (2009). REM sleep, prefrontal theta, and the consolidation of human emotional memory. *Cerebral Cortex, 19*(5), 1158-1166. doi.org/10.1093/cercor/bhn155

Payne, J. D., Schacter, D. L., Propper, R. E., Huang, L., Wamsley, E. J., Tucker, M. A., . . . Stickgold, R. (2009). The role of sleep in false memory formation. *Neurobiology of Learning and Memory, 92*(3), 327-334. doi:10.1016/j.nlm.2009.03.007

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

Pesta, B. J., Murphy, M. D., & Sanders, R. E. (2001). Are emotionally charged lures immune to false memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(2), 328-338. doi:10.1037/0278-7393.27.2.328

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rasch, B., & Born, J. (2007). Maintaining memories by reactivation. *Current Opinion in Neurobiology, 17*(6), 698-703. doi:10.1016/j.conb.2007.11.007

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

Ritchie, T. D., Batteson, T. J., Bohn, A., Crawford, M. T., Ferguson, G. V., Schrauf, R. W., . . . Walker, W. R. (2015). A pancultural perspective on the fading affect bias in autobiographical memory. *Memory, 23*(2), 278-290. doi:10.1080/09658211.2014.884138

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology, 21*(4), 803-814. doi:10.1037/0278-7393.21.4.803

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8*(3), 385-407. doi:10.3758/BF03196177

Ruci, L., Tomes, J. L., & Zelenski, J. M. (2009). Mood-congruent false memories in the DRM paradigm. *Cognition & Emotion, 23*(6), 1153-1165. doi:10.1080/02699930802355420

Schmidt, C., Collette, F., Cajochen, C., & Peigneux, P. (2007). A time to think: Circadian rhythms in human cognition. *Cognitive Neuropsychology, 24*(7), 755-789. doi:10.1080/02643290701754158

Shaw, J. J., & Monaghan, P. (2017). Lateralised sleep spindles relate to false memory generation. *Neuropsychologia, 107,* 60-67.

Sio, U. N., Monaghan, P., & Ormerod, T. (2013). Sleep on it, but only if it is difficult: Effects of sleep on problem solving. *Memory and Cognition*, *41*, 159-166. doi:10.3758/s13421-012-0256-7

Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition, 27*(3), 494-500. doi:10.3758/BF03211543

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature Cell Biology, 437*(7063), 1272-1278. doi:10.1038/nature04286

Straube, B. (2012). An overview of the neuro-cognitive processes involved in the encoding, consolidation, and retrieval of true and false memories. *Behavioral and Brain Functions, 8*, 35. doi: 10.1186/1744-9081-8-35

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: the mobilization-minimization hypothesis. *Psychological Bulletin, 110*(1), 67-85. doi:10.1037/0033-2909.110.1.67

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

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67*(6), 1176-1190. doi:10.1080/17470218.2013.850521

Vazquez, J., & Baghdoyan, H. A. (2001). Basal forebrain acetylcholine release during REM sleep is significantly greater than during waking. *American Journal of Physiology, 280*(2), R598-R601. doi: 10.1152/ajpregu.2001.280.2.R598

Walker, M. P., & Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron, 44*(1), 121-133. doi:10.1016/j.neuron.2004.08.031

Walker, W. R., Skowronski, J. J., & Thompson, C. P. (2003). Life is pleasant—and memory helps to keep it that way! *Review of General Psychology, 7*(2), 203-210. doi:10.1037/1089-2680.7.2.203

Walker, W. R., Vogl, R. J., & Thompson, C. P. (1997). Autobiographical memory: unpleasantness fades faster than pleasantness over time. *Applied Cognitive Psychology, 11*(5), 399-413.

Warriner, A., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(4), 1191-1207. doi:10.3758/s13428-012-0314-x

Zoellner, L. A., Foa, E. B., Brigidi, B. D., & Przeworski, A. (2000). Are trauma victims susceptible to "false memories"? *Journal of Abnormal Psychology, 109*(3), 517-524. doi:10.1037//0021-843x.109.3.517

Appendix

We ran the same general linear mixed effects models as in the main analysis, however we ran the analysis on all 111 participants with no exclusion criteria. Figure A1 shows the mean and standard error for each condition. The results revealed a significant main effect of word type, with increased performance accuracy for new than lure words (estimate = 1.993, SE = .265, z = 7.528, $p < .001$), and old than lure words (estimate = 1.358, SE = .223, z = 6.084, $p < .001$), again indicating that participants exhibited false memory effects for the lure words. There was also a main effect of emotionality, with neutral words responded to with overall increased accuracy than emotional words (estimate = .622, SE = .308, z = 2.016, $p = .044$). There was no significant main effect of valence, or group.

Similar to the main results, there was a significant word type by emotionality interaction for old versus lure words. Further exploration of the interaction for just the old words revealed greater accuracy for emotional than neutral words (estimate = -.463, SE = .186, z = -2.496, $p = .013$). Consistent with the main analysis, the lure words resulted in the opposite effect, with emotional words less accurate than neutral words (estimate = .500, SE = .156, z = 3.209, $p = .001$). Further analyses on emotional and neutral words separately indicated that, for emotional words, there was a significant difference between old and lure words (estimate = 1.139, SE = .145, z = 7.837, $p = 4.63 \times 10^{-15}$), with old words being accurately accepted significantly more than lure words were accurately rejected. However, for neutral words, there was no significant difference in performance between old and lure words, (estimate = .200, SE = .147, z = 1.357, $p = .175$).

We also found a significant 3-way interaction between word type, group, and valence (estimate = .766, SE = .385, z = 1.987, *p* = .047), as well as a marginally significant four-way interaction between word type, emotionality, valence, and group (estimate = -.887, SE = .528, z = -1.681, *p* = .093). The results revealed similar results to the main findings. First we analysed the sleep group, and found that for the positive group, there was a marginally significant difference in performance accuracy between emotional and neutral words for the old words (estimate = -.445, SE = .231, z = -1.927, *p* = .054), with positive words leading to greater accuracy, and a significant difference for lure words (estimate = .398, SE = .196, z = 2.032, *p* = .042), with increased false memories for positive than neutral words. Positive old words were responded to more accurately than positive lure words (estimate = .988, SE = .215, z = 4.602, *p* < .001), however neutral old and lure words did not significantly differ in response accuracy (estimate = .159, SE = .212, z = .752, *p* = .452). For those participants in the sleep group who were exposed to negative and neutral word lists, negative old words were remembered significantly better than neutral old words (estimate = -.571, SE = .264, z = -2.160, *p* = .031), and negative lure words led to increased false memories than neutral lure words (estimate = .617, SE = .296, z = 2.082, *p* = .037). Similar to the positive group, negative old words were remembered better than negative lure words (estimate = 1.409, SE = .286, z = 4.932, *p* < .001), and there was no significant difference in performance between neutral old and lure words, (estimate = .187, SE = .242, z = .773, *p* = .439).

For the wake group, analysis of the performance accuracy of participants in the positive and negative emotional stimuli groups separately revealed that positive old words were accurately responded to marginally significantly more than neutral old words (estimate = -.465, SE = .242, z = -1.925, *p* = .054), whereas neutral lure words were accurately responded to significantly more than positive lure words (estimate = .817, SE = .307, z = 2.664, *p* = .008). Furthermore, positive old words were accurately responded to significantly more than positive lure words, (estimate = 1.067, SE = .220, z = 4.855, *p* = 1.21x 10^{-6}), however there was no significant difference in performance between old and lure neutral words (estimate = -.103, SE = .294, z = -.349, *p* = .727). For those in the wake group who viewed negative and neutral word lists, there was no significant difference in performance between emotional and neutral words for either old (estimate = -.372, SE = .200, z = -1.860, *p* = .063) or lure words (estimate = .257, SE = .219, z = 1.175, *p* = .240). However negative old words were accurately responded to significantly more than negative lure words (estimate = 1.134, SE = .200, z = 5.668, *p* = 1.44 x 10^{-8}), and neutral old words were responded to significantly more accurately than neutral lure words, (estimate = .523, SE = .211, z = 2.481, *p* = .013).

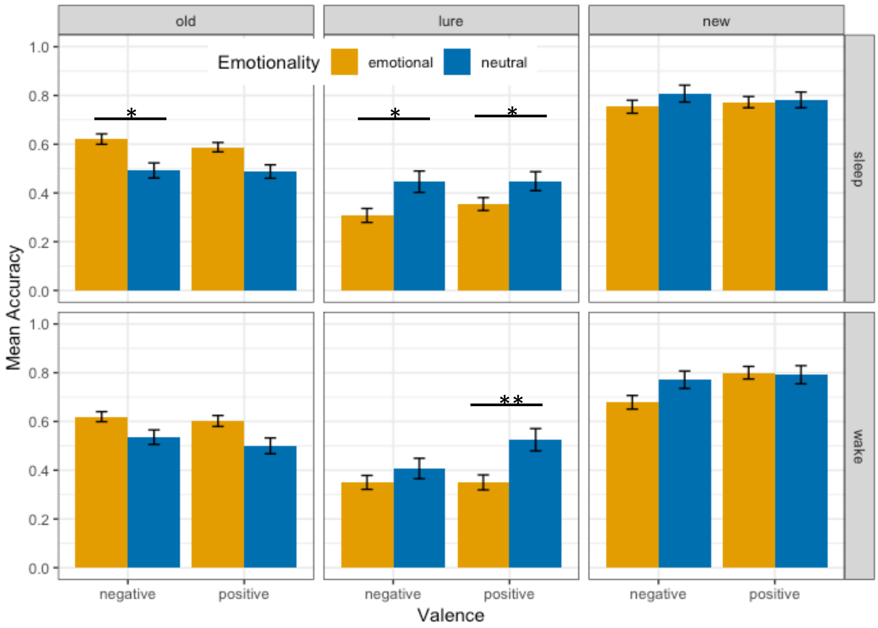


Figure 4.A1. Mean accuracy performance for old, lure, and new words of negative, positive, and neutral valence in the sleep and wake groups for all participants, error bars show ±1 SEM, **p* < .05, ***p* < .01.

Chapter 5: Sleep Spindles Relate to DRM Veridical but not False Memories.

The previous chapter explored the role of sleep in emotional memory processing, indicating that sleep-dependent memory consolidation supported negative emotional memory consolidation, but did not affect the formation of false memories. Furthermore, we found a specific influence of wakefulness on the consolidation of positive versus neutral memories. The following chapter sought to replicate these findings in a nap-paradigm, and also explore specific architecture that may be involved in the processing of emotional memories.

Statement of Author Contribution.

In the chapter entitled "Sleep Spindles Relate to DRM Veridical but not False Memories", the authors agree to the following contributions:

Chloe R. Newbury – 80% (Experimental design, data collection, data analysis, and writing)

Signed: C.Newbury Date: 25/04/2019

Professor Padraic Monaghan – 20% (Experimental design and review)

Signed: Date: 25/04/2019

Abstract

Sleep provides an overall boost to learning and memory, and emotional information may particularly benefit from offline memory consolidation during sleep. However, it is not yet understood what effect sleep has on the formation of false memories for emotional information, nor is it clear whether specific sleep stages are associated with memory performance for such thematically related emotive lists. We tested 60 participants on recognition performance for negative, positive, and neutral lists of thematically related words, after a short nap or wakefulness. We found that sleep increased veridical performance accuracy for emotional compared to neutral words. This increase in memory performance for emotional words was correlated with increased sleep spindle density during slow wave sleep. The results suggest that a short nap is therefore beneficial for improving veridical memory for emotional information, but does not affect the formation of false memories to the same extent.

1. Introduction

The beneficial effects of sleep on memory consolidation have been widely demonstrated (Marshall & Born, 2007; Rasch & Born, 2013; Walker & Stickgold, 2004). Models of memory consolidation propose that this sleep benefit is due to the reactivation of neuronal networks in the hippocampus, leading to the integration of memories to the long-term store in the neo-cortex, resulting in memory consolidation (Diekelmann & Born, 2010). In particular, evidence suggests that offline periods rich in slow wave sleep (SWS) are specifically important in the consolidation of declarative memories (Marshall & Born, 2007; van der Helm & Walker, 2009; Walker & Stickgold, 2004).

Non-rapid eye movement (NREM) sleep stage 2 and its subsequent sleep spindles have also been associated with increases in memory consolidation (Clemens, Fabó & Halász, 2006; Gais, Mölle, Helms & Born, 2002, Schmidt et al., 2006; Tamminen, Payne, Stickgold, Wamsley & Gaskell, 2010). Sleep spindles (11-15 Hz oscillations lasting up to 3s), are primarily associated with NREM stage 2, but also appear during SWS, where they have been found to correlate with hippocampal ripples (Siapas & Wilson, 1998). Sleep spindles have also been associated with triggering neural plasticity (Rosanova & Ulrich, 2005). Cox, Hofman, and Talamini (2012) suggest that although sleep spindles occur during both stage 2 and SWS sleep, sleep spindle density is more greatly correlated with memory during deep (SWS) sleep. Thus, sleep spindles may have different effects on memory dependent on sleep stages. Sleep spindles may therefore play a role in the reactivation and integration of memories from the hippocampal to neocortical systems, potentially supporting the consolidation of new memories into a pre-existing store.

Research examining memory for emotional information provides evidence for a particular enhancement in sleep-dependent memory consolidation compared to neutral information (Hu, Stylos-Allan & Walker, 2006; Payne, Stickgold, Swanberg & Kensinger, 2008; van der Helm & Walker, 2009; Walker, 2009). This beneficial effect of sleep specifically for emotional memory consolidation has been associated with increases in rapid-eye-movement (REM) sleep (Carr & Nielsen, 2015; Goldstein & Walker, 2014). Indeed, Nishida, Pearsall, Buckner and Walker (2009), using a nap study paradigm, found that participants who slept for 90mins after learning showed increased memory for negative over neutral stimuli, and this specifically correlated with time spent in REM sleep. Those who did not nap did not show this same enhancement for emotional memory. Specific characteristics of REM sleep have thus been linked to increases in emotional memory consolidation, in particular increases in levels of limbic and forebrain acetylcholine (Vazquez & Baghdoyan, 2001), which have been found to be important for emotional memory consolidation (McGaugh, 2004).

Although research indicates a role of REM sleep in emotional memory consolidation, studies also demonstrate that SWS may be associated with emotional memory processing (van der Helm & Walker, 2009; Walker & Stickgold, 2004), supporting an overall influence of SWS on memory (Marshall & Born, 2007; Walker, 2009). Further, sleep spindle density has also been implicated in emotional memory performance, with studies indicating a positive correlation between sleep spindles and emotional memories (Kaestner, Wixted, & Mednick, 2013). However, Cairney, Durrant, Jackson, and Lewis (2014) found a correlation between sleep spindles and slower response times for neutral stimuli, suggesting that increases in sleep spindles may be associated with the suppression of neutral information, thus indirectly impacting upon emotional memories.

Not only is sleep involved in supporting veridical memory consolidation, but it has also been found to affect memory for unseen, related information. Previous studies using the Deese-Roediger-McDermott (DRM) paradigm (Roediger & McDermott, 1995) expose participants to lists of semantically related words (eg., *bed, dream, tired, snooze, yawn,* etc.). During subsequent testing, participants often recall or recognise words that were not part of the original lists but were semantically similar, known as lure words. This is deemed the false memory effect (McDermott, 1996; Roediger & McDermott, 1995; Roediger, Watson, McDermott, & Gallo, 2001). Sleep has been found to enhance this false memory effect (Diekelmann, Born & Wagner, 2010; Monaghan, Shaw, Ashworth-Lord & Newbury, 2017; Payne et al., 2009), however the specific sleep architecture associated with false memories remains unclear. Pardilla-Delgado and Payne (2017) found a negative correlation between SWS and false memory recognition, but only in low performers, and a negative correlation between SWS and veridical memory performance has also been previously observed in DRM tests (Payne et al., 2009). Thus, one hypothesis is that SWS may be detrimental to the context-independent semantic processing necessary for the consolidation of thematically related lists within the DRM paradigm. Indeed, increases in SWS have been linked to reduced susceptibility to interference (Diekelmann, Büchel, Born, & Rasch, 2011), suggesting that increases in SWS may protect against false memories for unseen but semantically related information.

Conversely, SWS has been linked to the abstraction and integration of the gist of the memory (Tamminen et al., 2010), and sleep has also been proposed to support the reactivation of memories, integrating them with distantly related information (Cai, Mednick, Harrison, Kanady, & Mednick, 2009; Sio, Monaghan, & Ormerod, 2013), which drives an increase in false memories after sleep. Integration of newly learned information from hippocampal to neocortical systems is also found to occur much quicker if the information is linked to existing knowledge or schemas. Lewis and Durrant (2011) proposed a model that combined these findings, suggesting that reactivation of memories during SWS leads to strengthening of those memories, but also increases connections between memories and leads to schema formation and memory for the gist. This is termed the "information overlap to abstract" model (iOtA). If this is the case, then we would expect that increases in SWS may also lead to the enhancement of false memories.

Despite a range of evidence indicating greater memory consolidation and increasing false memories after sleep for neutral information, it is still unclear what effect sleep has on emotional memory for thematically related information within the DRM paradigm. Recent studies examining this suggest differing effects depending on the task. Whereas McKeon, Pace-Schott and Spencer (2012) found an increase in neutral than negative false memories after sleep in a test of recall, Newbury and Monaghan (submitted) found no effect of sleep on false recognition, but did find a specific effect of sleep on performance accuracy for negative emotional memory performance. Further research is therefore necessary to understand the way in which sleep prioritises the consolidation of negative, neutral, and positive information.

It thus remains unclear what effect sleep, and specific sleep stages, may have on emotional veridical and false memories after a daytime nap. The current study therefore extends the previous behavioural study of Newbury & Monaghan (submitted), that found an increase in veridical recognition memory for negative over neutral DRM words after sleep, but no effect on false memories, to investigate how sleep architecture relates to both veridical and false memories. Based on this previous behavioural study, we hypothesise that sleep will enhance veridical memory performance specifically for negative emotional words in comparison to neutral, whereas there will be no difference between positive and neutral memory performance. If SWS increases overall memory performance (Walker, 2009), then we would expect a positive correlation between SWS and overall memory accuracy. However, we may also expect a negative correlation between SWS and both veridical and false memories within the DRM paradigm, based on similar findings by Pardilla-Delgado and Payne (2017).

Furthermore, if sleep increases veridical memories, then we predict that sleep spindles will contribute to memory performance, with either a direct increase in sleep spindles for negative emotional memories (Kaestner et al., 2013), or an indirect effect on emotion, with sleep spindles suppressing memory for neutral word lists (Cairney et al., 2014). If sleep spindles support reactivation and integration of memories from hippocampal to neocortical systems (Siapas & Wilson, 1998), then we would also expect increases in sleep spindles to correlate with false memories. Indeed, Shaw & Monaghan (2017) in a nap study found that sleep spindles specifically in the right hemisphere correlated with enhanced false recognition using neutral DRM word lists.

2. Methods

2.1 Participants

There were 66 participants who took part in the study for either £24.50 (sleep group) or £21 (wake group). All participants completed a sleep diary for four nights prior to the study, to ensure that they maintained a normal sleep schedule, and were also required to avoid caffeine and alcohol the day of the study. All participants gave informed consent and were fully debriefed at the end of the study. Using the same criteria as Shaw and Monaghan (2017), we removed six participants from data analysis due to napping for less than 50 minutes. Sixty native speakers of English (41 females) with a mean age of 20.75 years (SD = 1.989, range = 18 to 30) were therefore included in the final analyses. Participants were randomly allocated to the sleep or wake group and to the negative or positive word lists condition. For the sleep groups, there were 17 participants in the negative condition (10 females), with a mean age of 20.88 years (SD = 1.677, range = 18 to 24), and 13 participants in the positive condition (9 females), with a mean age of 21.15 (SD = 1.748, range = 18 to 24). For the wake groups, there were 15 participants in the negative condition (13 female), with a mean age of 21.2 years (SD = 2.690, range = 18 to 30), and 15 participants in the positive condition (8 females), with a mean age of 19.8 years (SD = 1.223, range = 18 to 22).

2.2 Materials

We used the same materials as in Newbury and Monaghan (submitted). Negative and positive emotion DRM word lists and critical lures were taken from Brainerd, Holliday, Reyna, Yang, and Toglia (2010), who controlled for properties that are thought to affect false memory formation, including concreteness, meaning, and frequency of words (Roediger, Watson, McDermott, & Gallo, 2001). Neutral DRM lists and critical lures were taken from Stadler, Roediger, and McDermott (1999). There were five DRM lists of neutral words, ten negative lists, and ten positive lists.

During testing 60 words were used, two previously seen words from each DRM word list (total of 30), the critical lure associated with each DRM list (total of

15), and 15 unseen, unrelated words. Unrelated words were taken from lure words of unused DRM lists, and some of the unrelated emotional words were also taken from Kousta, Vinson, and Vigliocco (2009), who developed controlled sets of emotional and neutral word lists. Unrelated words were matched to DRM word lists in terms of valence, resulting in five unrelated neutral words, ten unrelated negative words and ten unrelated positive words. All words were presented in Courier new bold, black font, lower case and in 18-point.

2.3 Procedure

Participants first completed a sleep habits and caffeine/alcohol intake questionnaire, as well as the Self-Assessment Manikin (SAM; Bradley & Lang, 1994), the Stanford Sleepiness Scale (SSS; Hoddes, Dement & Zarcone, 1972), and a sleep diary to assess their sleeping patterns for the 4 nights prior to taking part in the study. Participants were then randomly allocated to the sleep or wake condition and sleep participants were attached to the polysomnography equipment. Participants were only told whether they would be napping after the training session.

Participants in both groups were then exposed to the 15 lists of 12 words. Participants were seated approximately 60cm from the computer screen and were instructed to attempt to remember the words that appeared on the screen. Participants were presented with a central fixation point for 500ms before words from one DRM list were presented one word at a time in the centre of the screen for 1500ms each. After each list, participants attempted to solve three simple maths problems that were presented for 10s each. The inclusion of maths problems acted as a distractor task, in order to prevent participants from rehearsing the word lists. Maths problems were presented in a random order for each participant, and each problem was only presented once throughout the training session. After participants had viewed the three maths problems, the fixation cross re-appeared and participants were given another DRM word list to remember. The presentation order of word lists was randomised, and the order in which each word in each list was presented was also randomised.

Sleep participants were given up to 120 min to nap, whereas those in the wake group watched an emotionally neutral movie with no verbal stimuli (a nature documentary with piano music played) for the same duration. Wake participants were monitored to ensure that they remained awake. After the two hours of sleep or wake, participants first had their sensors removed by the researcher and were given time to freshen up (~5 minutes), followed by a short break (~10 minutes) to reduce any potential effects of sleep inertia.

Participants then completed another SAM mood questionnaire and SSS questionnaire prior to the testing session. During the testing session, participants were asked to decide whether they recognised the word as one that was presented during the training session or not, by pressing ‘old’ or ‘new’ respectively on the keypad. Participants were presented with a central fixation cross for 500ms, before one of the test words was presented to the centre of the screen. Words were presented for 120ms each, and the presentation order of each word was randomised for each participant. Once participants had made their decision, they were then presented with another fixation point for 500ms, before another word appeared on the screen.

EEG recording and analysis. We recorded EEG (200Hz sampling rate) with an Embla N7000 system, with six scalp electrodes positioned at EEG sites F3, F4, C3, C4, O1, and O2, referenced to the contralateral mastoid (M1 and M2). We also monitored eye movements using two electrooculographic channels, and three chin electromyographic channels monitored muscle tone. All signals were verified at both the beginning and the end of participants' sleep time to have a connection impedance of <5kΩ. Noisy channels were removed from the final analysis. Sleep data were scored in 30s epochs in accordance with the standardised sleep scoring criteria of the American Association of Sleep Medicine (AASM). Spindle analysis was conducted during NREM sleep (stage 2 and SWS), with signals band-pass filtered (11-15 Hz) using a linear finite impulse response filter. We used an automatic spindle detection algorithm (Ferrarelli et al., 2007) to derive the number of discrete spindle events; amplitude fluctuations in the filtered time series exceeding a predetermined threshold counted as spindles. Thresholds were calculated relative to the mean channel amplitude (eight times the average amplitude). This algorithm has been widely used (e.g., Cairney et al., 2014; Tamminen et al., 2010, Shaw & Monaghan, 2017).

2.4 Data Analysis

We analysed both accuracy and response time (RT). Based on findings from Newbury & Monaghan (submitted), we predicted a four-way interaction between word type (old/lure), group (sleep/wake), emotionality (emotional/neutral), and valence (negative/positive), such that the sleep group would show increased performance accuracy for negative than neutral old words, and the wake group would show increased veridical and false memories for positive than neutral words. We further predicted a word type by emotionality interaction, such that emotional words led to increased veridical and false memories overall.

We analysed both response accuracy and response time (RT), using Linear Mixed-Effects Models (Baayen, 2008; Jaeger, 2008) in R (R Core Team, 2015). We took a hypothesis-driven approach to model building, thus, since we predicted a four-way interaction, we began with a maximal model including random effects of participants and items, as well as the four-way interaction between word type (old/lure/new), emotionality (emotional/neutral), valence (negative/positive), and group (sleep/wake). We removed any interaction or fixed effects if they did not significantly improve model fit. Exploratory analyses are reported separately.

For the EEG analysis, we performed a series of correlational analyses between each sleep stage and performance accuracy, and between sleep spindle density and performance accuracy. We took a hypothesis-driven approach to our correlation analyses. We predicted a correlation between SWS and both emotional veridical and false memories, as well as correlations between sleep spindle density and emotional veridical and false memories.

3. Results

3.1 Self-Assessment Manikin and Sleepiness Scale

Participants filled out a SAM mood questionnaire and Stanford Sleepiness Scale at two time points throughout the study, the first prior to taking part in the training session, and the second before the testing session. The SAM mood questionnaire was given to assess participants' ratings of valence (negative-positive), activation (calm-excited), and control (dominated-dominant). The SSS was given to assess participants' self-reported levels of sleepiness at both time periods. Mean scores of sleepiness, valence, activation, and control at both time points are reported in Table 1. The means show no difference in ratings between those in the sleep versus wake group, or between those exposed to negative and neutral versus positive and neutral word lists.

Table 5.1

Means and standard deviations of sleepiness score and SAM mood score prior to both sessions.

	Before Training				Before Testing			
	Sleep Group		Wake Group		Sleep Group		Wake Group	
	Negative Valence	Positive Valence	Negative Valence	Positive Valence	Negative Valence	Positive Valence	Negative Valence	Positive Valence
Mean(SD) sleepiness score	3.353(1.026)	3.077(0.616)	3.133(1.088)	2.933(0.854)	2.706(1.126)	2.692(0.822)	3.600(1.745)	3.867(1.360)
Mean(SD) valence mood score	6.412(1.033)	6.462(0.930)	6.400(1.357)	5.933(1.483)	6.412(1.141)	6.769(1.187)	5.467(1.708)	6.000(1.155)
Mean(SD) activation mood score	3.882(1.451)	3.846(1.293)	4.533(0.885)	4.733(1.770)	4.529(1.500)	4.692(1.589)	3.267(1.182)	3.267(2.082)
Mean(SD) control mood score	5.176(1.425)	5.538(1.152)	5.200(1.167)	5.467(1.669)	5.000(1.085)	5.308(1.589)	4.667(1.739)	4.733(1.570)

To ensure that mood and sleepiness had no effect on overall performance accuracy, we performed Generalized Linear Mixed-Effects Model analysis with recognition memory accuracy as the dependent variable, plus random effects of participants and items. Adding SAM valence score did not improve model fit compared to a model with only random effects at either time 1 ($\chi^2(1) = .834$, $p = .361$) or time 2 ($\chi^2(1) = 1.074$, $p = .300$). Activation score before both training ($\chi^2(1) = 1.136$, p = .287) and testing ($\chi^2(1) = .054$, $p = .816$) did not improve model fit, nor did control, ($\chi^2(1) = .008$, $p = .928$; $\chi^2(1) = 1.118$, $p = .290$). Similarly, sleepiness ratings before both training and testing did not significantly affect performance accuracy ($\chi^2(1) = .374$, $p = .541$; $\chi^2(1) = .269$, $p = .604$). This suggests no overall significant difference in performance accuracy dependent on either ratings of mood or ratings of sleepiness.

We then wanted to ensure that ratings of mood and sleepiness did not affect performance differently in the sleep versus wake groups, and those exposed to positive and neutral versus negative and neutral word lists. Adding the interaction term group (sleep/wake) by valence score did not improve model fit at either time 1 ($\chi^2(3) =$

1.362, $p = .715$) or time 2 ($\chi^2(3) = 1.098$, $p = .778$), nor did group by activation score ($\chi^2(3) = 2.041$, $p = .564$; $\chi^2(3) = .396$, $p = .941$), or group by control score ($\chi^2(3) = .260$, p = .967; $\chi^2(3) = 1.685$, $p = .640$). Ratings of sleepiness both before training and testing did not significantly effect performance accuracy in the sleep versus wake group ($\chi^2(3) = .771$, $p = .856$; $\chi^2(3) = .389$, $p = .942$). Similarly, adding the interaction term valence (positive and neutral/negative and neutral list exposure) by valence score at both times ($\chi^2(3) = 3.350$, $p = .341$; $\chi^2(3) = 4.168$, $p = .244$), activation score ($\chi^2(3) = 4.328$, $p = .228$; $\chi^2(3) = 3.433$, $p = .330$), control score ($\chi^2(3) = 2.517$, $p = .472$; $\chi^2(3) = 4.565$, $p = .207$), or sleepiness rating ($\chi^2(3) = 2.848$, $p = .416$; $\chi^2(3) = 2.826$, $p = .419$) did not improve model fit. Thus, participants' mood and sleepiness ratings in the wake versus sleep groups and negative versus positive valence groups did not significantly affect performance accuracy.

3.2 Accuracy

We performed a General Linear Mixed-Effects Model analysis (Baayen, 2008; Jaeger, 2008), to model the probability (log odds) of response accuracy, including random effects of subjects and items (words), and fixed effects of word type (old, lure, new), group (sleep, wake), emotionality (emotional, neutral) and valence (negative, positive), as well as the interactions between these factors. The full model is reported in Table 2, and Figure 1 shows mean performance accuracy for all conditions.

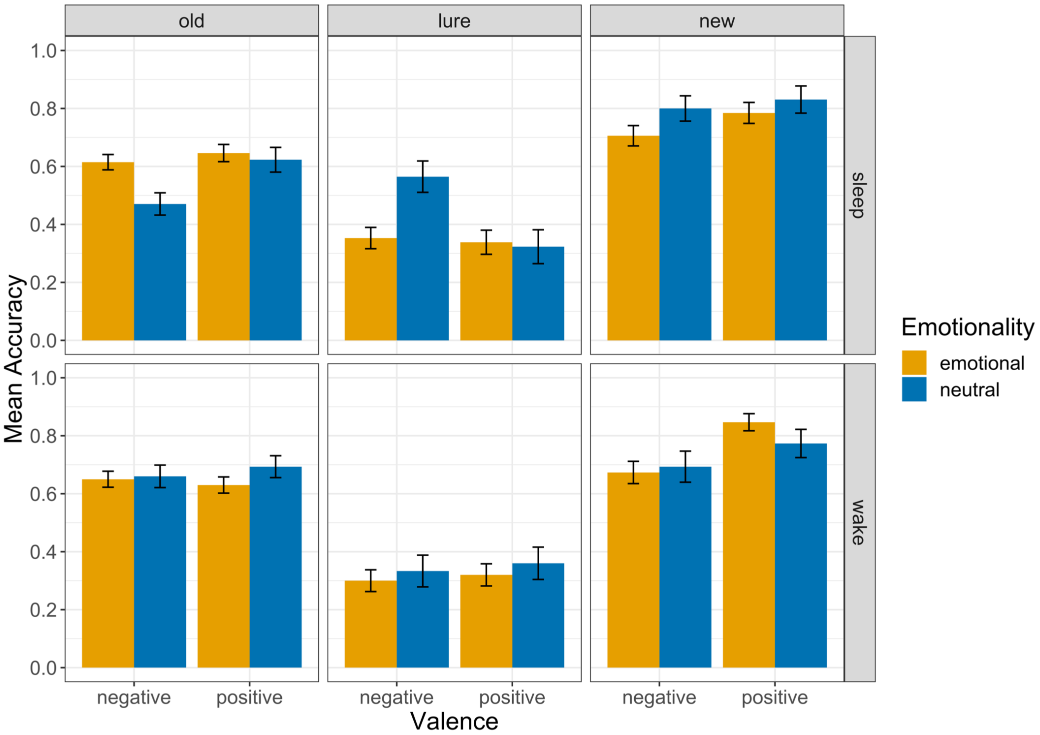


Figure 5.1. Mean performance accuracy for old, lure, and new words of negative, positive, and neutral valence in the sleep and wake groups, error bars show ±1 SEM.

Table 5.2

Summary of general linear mixed effects model of response accuracy, including all fixed effects and interaction terms.

	Estimated		Wald confidence intervals			
Fixed effects	coefficient	SE	2.50%	97.50%	z	Pr(>|z|)
(Intercept)	-0.645	0.219	-1.074	-0.215	-2.940	.003
Word type (new vs. lure)	1.560	0.312	0.950	2.171	5.009	<.001
Word type (old vs. lure)	1.137	0.266	0.617	1.658	4.280	<.001
Emotionality (neutral vs. emotional)	0.911	0.369	0.188	1.634	2.468	.014
Group (morning vs. evening)	-0.258	0.250	-0.748	0.232	-1.034	.301
Valence (positive vs. negative)	-0.054	0.323	-0.687	0.579	-0.168	.867
Word type (new vs. lure):emotionality interaction	-0.326	0.552	-1.409	0.757	-0.590	.555
Word type (old vs. lure):emotionality interaction	-1.524	0.451	-2.408	-0.641	-3.381	<.001
Word type (new vs. lure):group interaction	0.100	0.349	-0.585	0.785	0.286	.775
Word type (old vs. lure):group interaction	0.418	0.299	-0.169	1.005	1.395	.163
Emotionality:group interaction	-0.716	0.414	-1.527	0.095	-1.730	.084
Word type (new vs. lure):valence interaction	0.475	0.468	-0.441	1.391	1.016	.310
Word type (old vs. lure):valence interaction	0.194	0.392	-0.574	0.962	0.495	.621
Emotionality:valence interaction	-0.967	0.473	-1.894	-0.041	-2.047	.041
Group:valence interaction	0.171	0.362	-0.538	0.881	0.474	.636
Word type (new vs. lure):emotionality:group interaction	0.252	0.616	-0.956	1.459	0.408	.683
Word type (old vs. lure):emotionality:group interaction	1.358	0.504	0.370	2.346	2.694	.007
Word type (new vs. lure):emotionality:valence interaction	0.767	0.731	-0.666	2.200	1.050	.294
Word type (old vs. lure):emotionality:valence interaction	1.465	0.576	0.336	2.593	2.544	.011
Word type (new vs. lure):group:valence interaction	0.412	0.536	-0.638	1.463	0.770	.441
Word type (old vs. lure):group:valence interaction	-0.404	0.435	-1.257	0.449	-0.929	.353
Emotionality:group:valence interaction	0.970	0.607	-0.220	2.160	1.598	.110
Word type (new vs. lure):emotionality:group:valence interaction	-1.325	0.934	-3.156	0.506	-1.418	.156
Word type (old vs. lure):emotionality:group:valence interaction	-1.218	0.741	-2.671	0.234	-1.644	.100

Random effects	Name	Variance	Std. Dev
Subject effect on intercepts	(Intercept)	0.009	0.096
Item effect on intercepts	(Intercept)	0.199	0.446

	AIC	BIC	logLik	deviance
	4482.8	4643.7	-2215.4	4430.8

3600 observations, 60 participants, 60 items

Confirmatory Analyses

The results revealed a significant main effect of word type; previously seen old words were accurately accepted significantly more than unseen, related lure words were rejected, and unseen, unrelated new words were accurately rejected significantly more than lure words. There was also a main effect of emotionality, with neutral words responded to significantly more accurately than emotional words. There was no significant main effect of group or valence. This suggests that participants in the wake and sleep groups, and negative and positive valence groups, did not differ in their overall memory abilities.

The results also revealed a significant word type by emotionality interaction, as well as a three-way interaction between word type, emotionality, and sleep versus wake group (see Figure 2). Further General Linear Mixed-Effects Models for the wake and sleep group separately revealed that the two-way interaction between word type and emotionality was evident in the sleep group (estimate = -.879, SE = .315, z = -2.793, *p* = .005), but not the wake group (estimate = -.031, SE = .397, z = -.077, *p* = .938). We therefore explored the word type by emotionality interaction for the sleep group only. First, we investigated the effect of emotionality for old and lure words separately. For previously seen old words, neutral words were responded to with less accuracy than emotional words (estimate = -.402, SE = .177, z = -2.268, *p* = .023). For unseen, related lure words, there was no significant difference between emotional and neutral words (estimate = .526, SE = .320, z = 1.643, *p* = .100). We then investigated the effect of word type for emotional and neutral words separately. For emotional words, previously seen old words were accurately accepted significantly more than unseen, related lure words were rejected (estimate = 1.211, SE = .190, z = 6.387, *p* =

1.69×10^{-10}). For neutral words, there was no significant difference in performance accuracy between old and lure words (estimate = .307, SE = .201, z = 1.532, *p* = .126).

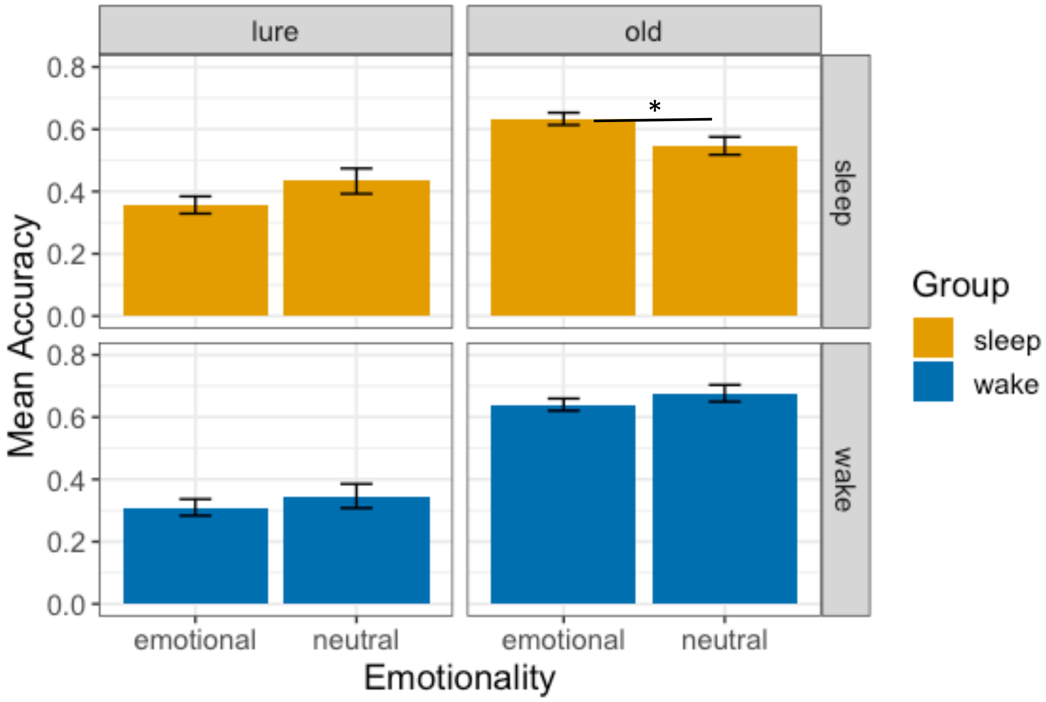


Figure 5.2. Mean performance accuracy for emotional and neutral old and lure words in the sleep and wake groups. Error bars show ±1 SEM, *p < .05.

Exploratory Analyses

The results further revealed a significant emotionality by valence interaction, and a three-way interaction between emotionality, valence and word type. Generalized Linear Mixed- Effects Models for negative and positive groups separately revealed a significant interaction between emotionality and word type for those who were exposed to negative and neutral word lists (estimate = -.906, SE = .403, z = -2.250, *p* = .025), but not for positive and neutral lists (estimate = .030, SE = .349, z = .085, *p* = .933). We therefore explored the emotionality by word type interaction in the negative group only. First, we investigated the effect of emotionality for old and lure words separately. For previously seen old words, there was no significant difference between emotional and neutral words (estimate = -.334, SE = .221, z = -1.509, *p* = .131), nor was there a significant difference for lure words (estimate = .605, SE = .390, z = 1.550, *p* = .121). Despite these non-significant results, the means indicate that for old words, emotional words were responded to with more accuracy than neutral words, whereas for lure words, emotional words were responded to with less accuracy than neutral words, indicating increased false recognition (See Figure 3). A lack of significance here may be explained by reduced power due to splitting of the data. We then investigated the effect of word type for emotional and neutral words separately. For both emotional and neutral words, previously seen old words were accurately accepted significantly more than unseen, related words were rejected (estimate = 1.359, SE = .276, z = 4.925, *p* = 8.42 x 10^{-7}; estimate = .414, SE = .196, z = 2.114, *p* = .035).

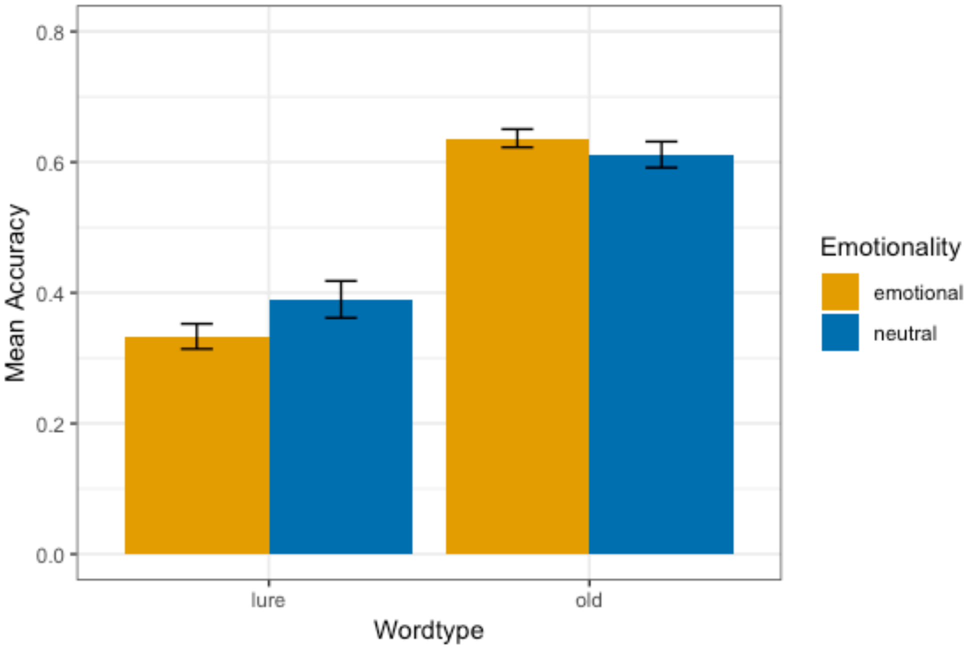


Figure 5.3. Mean performance accuracy of old and lure words for negative emotional versus neutral word lists.

3.3 Response Time

Response times (RT) for correct responses were analysed to establish whether any speed-accuracy trade-off effects influenced the results, as well as to test whether participants' speed of access to stored information differed between conditions. Responses that were slower than 2500ms and faster than 150ms were removed from the analysis (119 responses), resulting in the omission of 5.6% of the data, and leaving 2006 responses for analysis. Tests of normality revealed that the RT distribution was skewed, thus we reduced non-normality in the data by transforming latencies to $\log_{10}$(RT), based on the suggestions from Baayen, Feldman, and Schreuder (2006). We performed Linear Mixed-Effects Model analysis, predicting response time, with random effects of participants and items (words), and fixed effects of word type (old, lure, new), group (sleep, wake), emotionality (emotional, neutral), and valence (negative, positive), as well as the interaction between these factors. Based on suggestions from Baayen (2008), we accepted a coefficient as a significant predictor if $|t|>2$. The full model is reported in Table 3 and mean response time for each condition is reported in Figure 4.

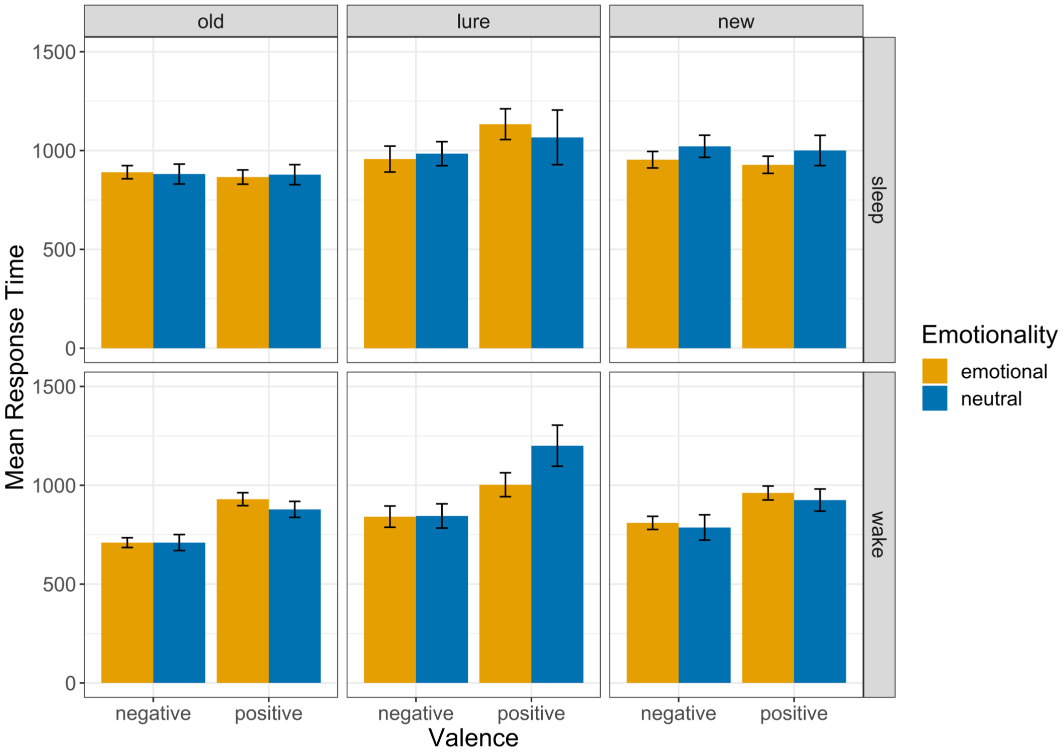


Figure 5.4. Mean performance accuracy for old, lure, and new words of negative, positive, and neutral valence in the sleep and wake groups, error bars show ±1 SEM.

Table 5.3

Summary of linear mixed effects model of response time, including all fixed effects and interaction terms.

Fixed effects	Estimated coefficient	SE	Wald confidence intervals 2.50%	97.50%	*t*
(Intercept)	2.924	0.033	2.860	2.989	89.23
Word type (new vs. lure)	0.013	0.028	-0.042	0.068	0.46
Word type (old vs. lure)	-0.026	0.026	-0.077	0.025	-0.99
Emotionality (neutral vs. emotional)	0.034	0.034	-0.033	0.101	1.00
Group (wake vs. sleep)	-0.049	0.048	-0.143	0.044	-1.03
Valence (positive vs. negative)	0.087	0.051	-0.012	0.187	1.73
Word type (new vs. lure):emotionality interaction	0.002	0.044	-0.084	0.087	0.04
Word type (old vs. lure):emotionality interaction	-0.034	0.041	-0.115	0.047	-0.82
Word type (new vs. lure):group interaction	-0.033	0.041	-0.113	0.047	-0.82
Word type (old vs. lure):group interaction	-0.038	0.038	-0.112	0.036	-1.00
Emotionality:group interaction	-0.022	0.053	-0.125	0.082	-0.41
Word type (new vs. lure):valence interaction	-0.085	0.043	-0.170	0.000	-1.96
Word type (old vs. lure):valence interaction	-0.094	0.041	-0.175	-0.015	-2.33
Emotionality:valence interaction	-0.091	0.059	-0.206	0.025	-1.53
Group:valence interaction	0.016	0.071	-0.122	0.154	0.23
Word type (new vs. lure):emotionality:group interaction	-0.032	0.065	-0.160	0.096	-0.50
Word type (old vs. lure):emotionality:group interaction	0.015	0.061	-0.105	0.134	0.24
Word type (new vs. lure):emotionality:valence interaction	0.063	0.071	-0.076	0.202	0.89
Word type (old vs. lure):emotionality:valence interaction	0.115	0.068	-0.017	0.248	1.70
Word type (new vs. lure):group:valence interaction	0.079	0.059	-0.037	0.195	1.33
Word type (old vs. lure):group:valence interaction	0.106	0.060	-0.004	0.216	1.90
Emotionality:group:valence interaction	0.148	0.082	-0.013	0.309	1.80
Word type (new vs. lure):emotionality:group:valence interaction	-0.123	0.099	-0.317	0.071	-1.24
Word type (old vs. lure):emotionality:group:valence interaction	-0.180	0.093	-0.363	0.003	-1.93

Random effects	Name	Variance	Std. Dev
Subject effect on intercepts	(Intercept)	0.009	0.164
Item effect on intercepts	(Intercept)	<.001	0.017

AIC	BIC	logLik	deviance
-1347.6	-1196.3	700.8	-1401.6

2006 observations, 60 participants, 60 items

The results revealed a significant word type by valence interaction for old versus lure words. First, we looked at positive and negative groups separately, which revealed that old words were accurately accepted quicker than lure words were rejected for both positive and neutral words (estimate = -.085, SE = .020, *t* = -4.290) and negative and neutral words (estimate = -.051, SE = .016, *t* = -3.230). We then conducted analyses on old and lure words separately, with results indicating that for previously seen old words, there was no significant difference in RT between negative and positive valence groups (estimate = .054, SE = .029, *t* = 1.84). For unseen, related lure words, those who viewed positive and neutral words responded significantly slower than those who were exposed to negative and neutral words (estimate = .081, SE = .033, *t* = 2.480). There were no other main effects or interactions.

3.4 Sleep Stage Analysis

We measured total sleep time, and time spend in sleep stage 1, stage 2, SWS and REM sleep (see Table 4 for mean durations). We correlated the proportion of time in each stage with memory accuracy for old, lure, and new words, as well as memory accuracy for emotional and neutral lists. REM sleep was removed from the analysis since only one participant entered this stage. We employed the Bonferroni correction to control for multiple comparisons. There were no significant correlations between proportion of total sleep time in any sleep stage and memory accuracy (see Table 5). Since the behavioural results revealed a significant interaction between old versus lure words and emotionality in the sleep group, we conducted further correlations between old and lure emotional words and old and lure neutral words. Again, all correlations were non-significant.

Table 5.4

Mean total sleep time, wake time, and sleep onset in minutes, and proportion of time spend in stage 1, stage 2, SWS, and REM sleep, ±1 SEM.

Sleep Descriptives	Minutes ± 1 SEM
Total sleep time	84.783 ± 3.448
Average wake after sleep onset	23.300 ± 2.735
Average sleep onset latency	15.817 ± 2.176

Sleep Stages	Proportion ± 1 SEM
Stage 1	0.148 ± 0.016
Stage 2	0.386 ± 0.032
SWS	0.149 ± 0.020
REM	0.006 ± 0.006

Table 5.5

Correlation coefficients between recognition accuracy and proportion of time in each sleep stage.

	Seen-old word recognition	Unseen-lure word recognition	Unseen-new word recognition	Emotional words	Neutral words
Stage 1	0.154	0.015	0.177	0.253	0.101
Stage 2	0.157	-0.320	-0.158	-0.034	-0.396
SWS	-0.025	0.010	-0.221	-0.123	-0.128

3.5 Sleep Spindle Analysis

We analysed whether there were any correlations between memory accuracy and sleep spindle density. We analysed spindle-per-minute density averaged over all electrode sites, as well as split by C3 and C4 combined electrode sites (C3-C4), or F3 and F4 combined electrode sites (F3-F4). Since sleep spindles have been found to have distinct memory effects depending on the different sleep stages (Cox et al., 2012), we measured sleep spindles in stage 2 and SWS separately.

3.5.1 Stage 2 sleep spindles

Confirmatory Analyses. Since we found different behavioural effects for emotional versus neutral word lists for old and lure words in the sleep group, we correlated sleep spindle density with old and lure performance for emotional and neutral words separately. First, we investigated overall sleep spindle density, and then looked at spindle density in C3-C4 and F3-F4 electrode sites separately. All correlations were non-significant.

Exploratory Analyses. We also correlated sleep spindle density with overall memory accuracy for old, lure, and new words, as well as memory accuracy for emotional and neutral lists. Again, we found no significant correlations.

3.5.2 SWS sleep spindles

Confirmatory Analyses. Again we investigated the specific correlations between spindle density and memory accuracy for old and lure words split by emotionality. We found a significant correlation between overall spindle density and memory accuracy for old emotional words ($r = .540$, $p = .030$ corrected, see Figure 6). Analysing this correlation for C3-C4 and F3-F4 electrode sites separately, the significant correlation between spindle density and old emotional words was found specifically within F3-F4 electrode sites ($r = .538$, $p = .007$ uncorrected).

Exploratory Analyses. We then investigated overall correlations for each word type and emotionality. There were no significant correlations between sleep spindle density and memory accuracy for old, lure, and new words, or for memory accuracy for emotional and neutral lists, either for overall spindle density, or by C3-C4 and F3-F4 electrode sites.

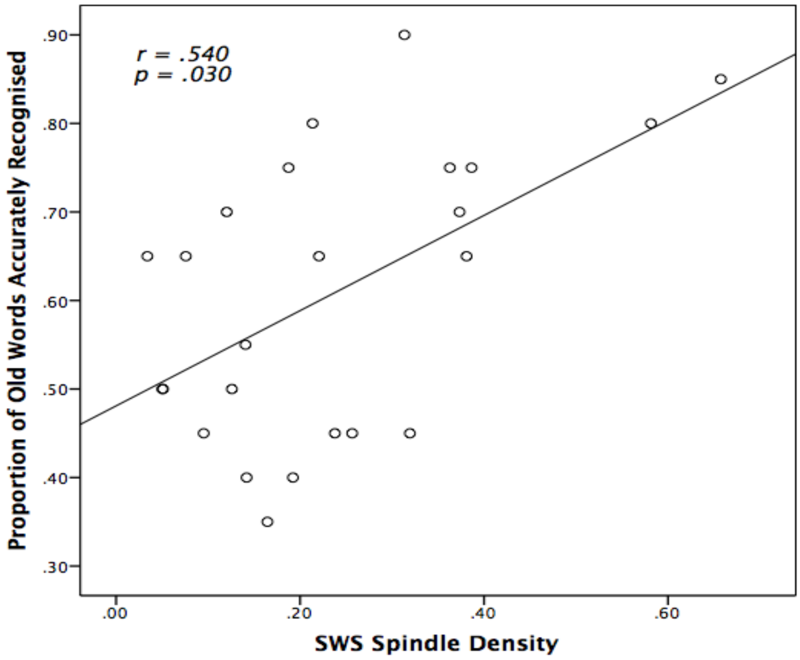


Figure 5.6. Correlation between SWS spindle density and seen-old word recognition accuracy.

4. Discussion

A main aim of this study was to investigate the role that sleep plays in the reactivation and integration of emotional memories into the long-term store. We followed the same methodology as Newbury and Monaghan (submitted), who found increased performance accuracy for negative than neutral DRM lists after sleep and increased veridical and false memories for positive lists after wake. We replicated the central aspect of this finding, with a short nap leading to increases in performance accuracy for emotional than neutral word lists, although we did not find a specific effect of sleep on negative emotional information, but rather an overall increase in performance for emotional lists. We suggest that this lack of an effect of negative versus positive lists may be due to the shorter time period that participants were given to sleep in the current study. Mean total sleep time in the current study was 85 minutes, whereas in the previous study, participants slept for at least six hours. Since evidence suggests that REM sleep in particular increases spreading activation (Cai et al., 2009), we may be more likely to see performance differences between negative and positive word lists after a period of overnight sleep rich in REM. Evidence does suggest that negative emotion in particular is associated with REM sleep (Nishida et al., 2009), thus it could be that although sleep aids in overall memory consolidation for emotional information, over the period of a night of sleep containing REM sleep, the reactivation and integration of memories becomes more specific to negative emotional memories.

A second aim of the current study was to understand what specific sleep architecture is involved in both veridical and false memory consolidation for thematically related emotional word lists. We found an association between sleep spindle density during SWS and increased accuracy for emotional memories. Sleep spindles have been previously correlated with hippocampal ripples (Siapas & Wilson, 1998) and neural plasticity (Rosanova & Ulrich, 2005), and thus the present study adds to literature suggesting an importance of sleep spindles in the integration of memories from hippocampal and neocortical systems to the neo-cortex. Additionally, previous findings suggest a positive correlation between pharmacologically induced sleep spindles and emotional memories (Kaestner et al., 2013), providing further support for our findings of increased SWS spindles relating to emotional memory accuracy.

Our results further support previous research indicating a specific involvement of SWS in memory consolidation, and in particular emotional memory processing (Marshall & Born, 2007; Tamminen et al., 2010; van der Helm & Walker, 2009; Walker, 2009; Walker & Stickgold, 2004). The iOtA model (Lewis & Durrant, 2011) proposes that reactivation of memories during SWS leads to both strengthening of memories, as well as increasing semantic associations that lead to memory for the gist. The current study partly supports this model, indicating an increase in emotional veridical memories related to SWS, however we did not find increases in false recognition relating to SWS. The lack of an effect of SWS spindle density on emotional false memories may be explained by two possibilities. First, since the behavioural results indicate no effect of sleep on emotional false memories, it could be that sleep plays no role in the formation of false memories in tests of recognition. Previous studies of neutral false memories after sleep indicate either no effect of sleep on false recognition or a reduction in the false memory effect (Diekelmann, Londolt, Lahl, Born, & Wagner, 2008; Fenn, Gallo, Margoliash, Roediger, & Nusbaum, 2009) and a recent meta-analysis provides further evidence for a lack of an effect of sleep on false recognition memory (Newbury & Monaghan, 2018). This difference between tests of recall and recognition within the DRM paradigm can be explained by the activation/monitoring framework (Collins & Loftus, 1975), which proposes that during tests of recognition, participants are able to activate monitoring cues when words are presented, allowing for the rejection of thematically related words that were unseen. During tests of recall these monitoring cues are not available, and thus participants find it more difficult to suppress unseen but related words, leading to increased false recall.

A second possibility is that no effect of SWS spindle density on false recognition was found due to the short amount of time that participants slept. Participants spent on average 18 minutes in SWS, and thus after a longer period of sleep we may find greater effects of SWS on emotional veridical and false memories. However, in contrast, a longer period of sleep could also lead to a negative correlation between SWS and the formation of emotional false memories, based on previous evidence (Payne et al., 2009). Consequently, how sleep and sleep architecture relates to false memories, and in particular emotional false memories, is not yet entirely resolved.

A potential limitation of these results is that multiple significance tests were conducted to investigate interaction effects. Thus, it is possible that our findings may be subjected to type one errors, due to increased comparisons with a low sample size. Future research should focus on replicating the effects of sleep on emotional veridical and false memories, and further investigate the role of specific sleep stages on memory performance at a larger scale.

In summary, our findings indicate a behavioural effect of sleep on increased accuracy for emotional compared to neutral information, and this effect is associated with increasing SWS spindle density. The findings of this study therefore add to a growing literature that highlights an effect of sleep on emotional memory performance, and supports an influence of SWS and subsequent sleep spindles in offline memory consolidation. This research is of particular importance in various fields of research, for example when maximising performance accuracy when recalling emotional events such as during eye-witness testimony, or minimizing the impact of negative emotional events, such as in those with post-traumatic stress disorder or depression, who have been found to be more susceptible to emotional memory effects (e.g., Jelinek, Hottenrott, Randjbar, Peter, & Mortiz, 2009; Brennen, Dybdahl, & Kapidžić, 2007). Further research focusing on how sleep architecture, and in particular sleep spindles, might relate to both emotional veridical and false memories during longer periods of overnight sleep will allow us to gain a greater understanding as to the mechanisms of sleep affecting the integration of emotional memories into the long-term store, and the subsequent retrieval of these memories through either recall or recognition.

References

Baayen, R. H. (2008). *Analyzing linguistic data [electronic resource] : a practical introduction to statistics using R*. Cambridge: Cambridge : Cambridge University Press.

Brainerd, C. J., Holliday, R., Reyna, V., Yang, Y., & Toglia, M. (2010). Developmental reversals in false memory: Effects of emotional valence and arousal. *Journal of Experimental Child Psychology, 107*(2), 137-154. doi:10.1016/j.jecp.2010.04.013

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49-59. doi:10.1016/0005-7916(94)90063-9

Brennen, T., Dybdahl, R., & Kapidžić, A. (2007). Trauma-related and neutral false memories in war-induced Posttraumatic Stress Disorder. *Consciousness and Cognition, 16*(4), 877-885. doi:10.1016/j.concog.2006.06.012

Cai, D., Mednick, S. A., Harrison, E, M., Kanady, J. C., & Mednick, S. C. (2009). REM, not incubation, improves creativity by priming associative networks. *Proceedings of the National Academy of Sciences, 106*, 10130–10134. doi:10.1073/pnas.0900271106

Cairney, S. A., Durrant, S. J., Jackson, R., & Lewis, P. A. (2014). Sleep spindles provide indirect support to the consolidation of emotional encoding contexts. *Neuropsychologia, 63*, 285-292. doi:10.1016/j.neuropsychologia.2014.09.016

Carr, M., & Nielsen, T. (2015). Morning rapid eye movement sleep naps facilitate broad access to emotional semantic networks. *Sleep, 38*(3), 433-443. doi:10.5665/sleep.4504

Clemens, Fabó, & Halász. (2006). Twenty-four hours retention of visuospatial memory correlates with the number of parietal sleep spindles. *Neuroscience Letters, 403*(1), 52-56. doi:10.1016/j.neulet.2006.04.035

Collins, A., & Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407-428.

Cox, R., Hofman, W. F., & Talamini, L. M. (2012). Involvement of spindles in memory consolidation is slow wave sleep-specific. *Learning & Memory, 19*(7), 264-267. doi:10.1101/lm.026252.112

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience, 11*(2), 114-126. doi:10.1038/nrn2762

Diekelmann, S., Born, J., & Wagner, U. (2010). Sleep enhances false memories depending on general memory performance. *Behavioural Brain Research, 208*(2), 425-429. doi:10.1016/j.bbr.2009.12.021

Diekelmann, S., Büchel, C., Born, J., & Rasch, B. (2011). Labile or stable: Opposing consequences for memory when reactivated during waking and sleep. *Nature Neuroscience, 14*(3), 381-386. doi:10.1038/nn.2744

Diekelmann, S., Landolt, H., Lahl, O., Born, J., & Wagner, U. (2008). Sleep loss produces false memories. *PloS one, 3*(10), e3512-e3512. doi:10.1371/journal.pone.0003512

Fenn, K. M., Gallo, D. A., Margoliash, D., Roediger, H. L., & Nusbaum, H. C. (2009). Reduced false memory after sleep. *Learning & Memory, 16*(9), 509-513. doi:10.1101/lm.1500808

Ferrarelli, F., Huber, R., Peterson, M. J., Massimini, M., Murphy, M., Riedner, B. A., Watson, A., Bria, P., Tononi, G. (2007). Reduced sleep spindle activity in schizophrenia patients. *American Journal of Psychiatry, 164,* 483–492. doi:10.1176/appi.ajp.164.3.483

Gais, S., Mölle, M., Helms, K., & Born, J. (2002). Learning-dependent increases in sleep spindle density. *The Journal of Neuroscience, 22*(15), 6830-6834. doi:10.1523/JNEUROSCI.22-15-06830.2002

Goldstein, A. N., & Walker, M. P. (2014). The role of sleep in emotional brain function. *Annual Review of Clinical Psychology, 10*, 679-708. doi:10.1146/annurev-clinpsy-032813-153716

Hoddes, R., Dement, W., & Zarcone, V. (1972). The development and use of the Stanford Sleepiness Scale (SSS). *Psychophysiology, 9*, 150.

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

Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (Transformation or Not) and towards Logit Mixed Models. *Journal of Memory and Language, 59*(4), 434-446. doi:10.1016/j.jml.2007.11.007

Jelinek, L., Hottenrott, B., Randjbar, S., Peters, M. J., & Moritz, S. (2009). Visual false memories in post-traumatic stress disorder (PTSD). *Journal of Behavior Therapy and Experimental Psychiatry, 40*(2), 374-383. doi:10.1016/j.jbtep.2009.02.003

Kaestner, E. J., Wixted, J. T., & Mednick, S. C. (2013). Pharmacologically Increasing Sleep Spindles Enhances Recognition for Negative and High-arousal Memories. *Journal of Cognitive Neuroscience*, *25*(10), 1597-1610. doi:10.1162/jocn_a_00433

Kousta, S. T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition, 112*(3), 473-481. doi:10.1016/j.cognition.2009.06.007

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

Marshall, L., & Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends in Cognitive Sciences, 11*(10), 442-450. doi:10.1016/j.tics.2007.09.001

McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language, 35*(2), 212-230. doi:10.1006/jmla.1996.0012

McKeon, S., Pace-Schott, E. F., & Spencer, R. M. (2012). Interaction of sleep and emotional content on the production of false memories. *PloS one, 7*(11), e49353. doi:10.1371/journal.pone.0049353

McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience, 27*(1), 1-28. doi:10.1146/annurev.neuro.27.070203.144157

Monaghan, P., Shaw, J. J., Ashworth-Lord, A., & Newbury, C. R. (2017). Hemispheric processing of memory is affected by sleep. *Brain and Language, 167*, 36-43. doi:10.1016/j.bandl.2016.05.003

Newbury, C. R., & Monaghan, P. (2018). *Negative but not positive emotional memories are enhanced by sleep*. Manuscript submitted for review.

Newbury, C. R., & Monaghan, P. (2018). When does sleep affect veridical and false memory consolidation?: A meta-analysis. *Psychonomic Bulletin & Review*, 1-14.

Nishida, M., Pearsall, J., Buckner, R. L., & Walker, M. P. (2009). REM sleep, prefrontal theta, and the consolidation of human emotional memory. *Cerebral Cortex, 19*(5), 1158-1166. doi.org/10.1093/cercor/bhn155

Pardilla-Delgado, & Payne. (2017). The impact of sleep on true and false memory across long delays. *Neurobiology of Learning and Memory,137*, 123-133. doi:10.1016/j.nlm.2016.11.016

Payne, J. D., Schacter, D. L., Propper, R. E., Huang, L., Wamsley, E. J., Tucker, M. A., . . . Stickgold, R. (2009). The role of sleep in false memory formation. *Neurobiology of Learning and Memory, 92*(3), 327-334. doi:10.1016/j.nlm.2009.03.007

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

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

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

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology, 21*(4), 803-814. doi:10.1037/0278-7393.21.4.803

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8*(3), 385-407. doi:10.3758/BF03196177

Rosanova, M., & Ulrich, D. (2005). Pattern-specific associative long-term potentiation induced by a sleep spindle-related spike train. *The Journal of Neuroscience, 25*(41), 9398-405. doi:10.1523/JNEUROSCI.2149-05.2005

Schmidt, C., Peigneux, P., Muto, V., Schenkel, M., Knoblauch, V., Münch, M., . . . Cajochen, C. (2006). Encoding difficulty promotes postlearning changes in sleep spindle activity during napping. *The Journal of Neuroscience, 26*(35), 8976-8982. doi:10.1523/JNEUROSCI.2464-06.2006

Shaw, J. J., & Monaghan, P. (2017). Lateralised sleep spindles relate to false memory generation. *Neuropsychologia, 107,* 60-67. doi:10.1016/j.neuropsychologia.2017.11.002

Siapas, A. G., & Wilson, M. A. (1998). Coordinated Interactions between Hippocampal Ripples and Cortical Spindles during Slow-Wave Sleep. *Neuron, 21*(5), 1123-1128. doi:10.1016/S0896-6273(00)80629-7

Sio, U. N., Monaghan, P., & Ormerod, T. (2013). Sleep on it, but only if it is difficult: Effects of sleep on problem solving. *Memory and Cognition*, *41*, 159-166. doi:10.3758/s13421-012-0256-7

Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition, 27*(3), 494-500. doi:10.3758/BF03211543

Tamminen, J., Payne, J., Stickgold, R., Wamsley, E., & Gaskell, M. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *The Journal of Neuroscience, 30*(43), 14356-14360. doi:10.1523/JNEUROSCI.3028-10.2010 van der Helm, E., Walker, M. P. (2009). Overnight therapy? The role of sleep in emotional brain processing. *Psychological Bulletin, 135*(5), 731-748. doi:10.1037/a0016570

Vazquez, J., & Baghdoyan, H. A. (2001). Basal forebrain acetylcholine release during REM sleep is significantly greater than during waking. *American Journal of Physiology, 280*(2), R598-R601. doi:10.1152/ajpregu.2001.280.2.R598

Walker, M. P. (2009). The Role of Sleep in Cognition and Emotion. *Annals of the New York Academy of Sciences, 11561*(1), 168-197. doi:10.1111/j.1749-6632.2009.04416.x

Walker, M. P., & Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron, 44*(1), 121-133. doi:10.1016/j.neuron.2004.08.031

Chapter 6: A Role of Sleep in the Consolidation of Emotional Metaphorical Word Pairs.

The previous two chapters explored the role of sleep in both veridical and false memories within the DRM paradigm. The findings provide support for the selective enhancement of emotional memories that is related to sleep-dependent memory consolidation. However, whether sleep boosts both negative and positive emotion or is specific to negative emotionality may be dependent on the duration of sleep.

The following study examined whether an effect of sleep on emotional memory processing can be extended to another type of stimuli. We investigated the role of sleep in the processing of figurative language, and whether the emotionality of metaphorical word pairs influenced the learning of and memory for novel metaphorical stimuli.

Statement of Author Contribution.

In the chapter entitled “A Role of Sleep in the Consolidation of Emotional Metaphorical Word Pairs”, the authors agree to the following contributions:

Chloe R. Newbury – 70% (Experimental design, data collection, data analysis, and writing)

Signed: C Newbury Date: 25/04/2019

Dr. Francesca Citron – 15% (Experimental design and review)

Signed: F. Citron Date: 25/04/2019

Professor Padraic Monaghan – 15% (Experimental design and review)

Signed: P Monaghan Date: 25/04/2019

Abstract

Memory for emotional information is greater than that of non-emotional information, and emotional memories are further enhanced by sleep-related consolidation. Inconsistencies within the literature make it unclear whether this boost to emotional memories is due to effects of emotional arousal or of valence. Figurative expressions in particular are often rated as higher in both valence and arousal than their literal counterparts and are used frequently in everyday language. However, little research has investigated the effect of emotionality on the learning of such metaphorical language, nor have studies examined the differential effect of sleep on emotionally negative, positive, and neutral language. In the current study, participants learned conventional (e.g., *sunny disposition*) and novel (e.g., *cloudy disposition*) metaphorical word pairs and definitions varying in valence. After a 12-hour period of sleep or wake, participants were tested on their recall of definitions and recognition of word pairs. We found higher arousal ratings were related to increased recall and recognition performance. Furthermore, sleep particularly increased accurate recognition of word pairs of both negative and positive valence compared to wake. The results provide support for a benefit of sleep in the consolidation of both negative and positive emotional memories, and indicates a specific influence of arousal in the processing of figurative language.

1. Introduction

A large body of literature suggests that memory for emotional information is much more accurate than that of non-emotional information (Adelman & Estes, 2013; Kensinger & Corkin, 2004; McGaugh, 2004). Dimensional models of affect suggest differences between emotional valence, which describes how positive or negative a stimulus is, and arousal, which can range from calm to exciting or agitating (LaBar & Cabeza, 2006; Reisenzein, 1994). Although an overall effect of emotion on memory processing is evident, inconsistencies emerge between effects of arousal and effects of valence. Words high in negative valence and high in arousal have been found to lead to the greatest increases in memory performance (Kensinger & Corkin, 2003), supporting suggestions that stimuli high in negative valence have been found to correlate with higher ratings of arousal (Ito, Cacciopo, & Lang, 1998). In contrast, Adelmann and Estes (2013) report significant increases in memory performance for words of both negative and positive valence, however no effect of arousal was found. Thus, although differences in memory performance for stimuli varying in emotionality emerge, specific effects of valence and arousal remain unclear.

A range of evidence indicates a role of sleep in both learning and memory (Born, Rasch, & Gais, 2006; Rasch & Born, 2013). Research indicates that a period of sleep after learning increases memory for word pairs compared to wakefulness (Plihal & Born, 1997; Wilson, Baran, Pace-Schott, Ivry, & Spencer, 2012), and integration of novel word pairs into the long-term store has also been found to occur after a period of sleep (Dumay & Gaskell, 2007). The Active Systems Consolidation hypothesis (Diekelmann & Born, 2010; Marshall & Born, 2007; Rasch & Born, 2013) suggests that this boost to learning and memory during offline periods of sleep is due to the reactivation and integration of these memories from hippocampal systems where they are initially encoded, into the long-term store of the neocortex.

Sleep has been found to enhance the processing of emotional information in particular (van der Helm & Walker, 2009; Walker, 2009). Payne, Stickgold, Swanberg, and Kensinger (2008) found that sleep increased memory for negative parts of a scene, but not for background information, and findings indicate a similar effect when investigating memory performance for positive versus neutral stimuli (Chambers & Payne, 2014). Similarly, Hu, Stylos-Allen, and Walker (2006) found evidence of greater memory for emotionally arousing than non-emotional images after a period of sleep compared to wake, and Nishida, Pearsall, Buckner, and Walker (2009) found a benefit of sleep for remembering negative, arousing stimuli, but not for neutral, non-arousing stimuli. This increase in memory for emotional stimuli was found to correlate with rapid-eye-movement (REM) sleep, and a range of research supports a specific role of REM sleep in emotional memory consolidation (Groch, Wilhelm, Diekelmann, & Born, 2012; Wagner, Gais, & Born, 2001).

Despite evidence of increased performance accuracy after sleep for emotional stimuli, findings do not disentangle effects of arousal from effects of valence. Although findings suggest an effect of both negative and positive valence (Chambers & Payne, 2014; Payne et al., 2008), no studies have investigated the effect of sleep on memory for negative compared to positive valence, nor have they examined effects of ratings of arousal versus ratings of valence. Given the potential differences between valence and arousal on memory performance without sleep (e.g., Adelmann & Estes, 2013; Kensinger & Corkin, 2003), it is important to understand whether this sleep-related benefit for emotional memory consolidation is due to effects of arousal independent of valence, or due to specific differences between stimuli of positive and negative valence.

A key type of emotional stimulus is figurative language, where the meaning of the stimulus is not directly available from the words it comprises, but is learned by convention. Figurative language is prevalent and common in language (Carter, 2004). The Conceptual Metaphor Theory (Lakoff & Johnson, 1980) suggests this ease of use of metaphors in particular may be due to a mapping between metaphorical expressions that are often more concrete, for example *bright*, and an abstract, literal expression, e.g., *smart*. Thus, metaphorical expressions are useful for conceptualising more abstract experiences and creating concrete concepts that may be easier to understand. In particular, previous findings suggest that metaphorical language is used much more frequently to describe events that are high in emotional intensity (Fainsilber & Ortony, 1987), and so it is a fundamental aspect of human emotional expression. More recently, Citron and Goldberg (2014) found that conventional metaphors conveying taste were more emotionally engaging than literal sentences, and Citron, Güsten, Michaelis, and Goldberg (2016) found similar results for longer passages containing metaphorical language, not restricted to taste. These neuroimaging results indicate increased activation of the left hippocampus and amygdala when processing metaphorical as opposed to literal sentences. Greater activity in the amygdala has been reported to support emotional memory processing (Cahill et al., 1996; McGaugh, 2004), and the encoding of emotional memories has been correlated with co-activation of both the amygdala and hippocampus (for a review see Walker, 2009).

Despite a range of research surrounding the processing of figurative language, there has been little investigation into how such non-literal expressions are learned and remembered. Evidence of increased activation in a number of brain regions (for a review see Citron & Goldberg, 2014) suggests that greater processing resources are involved in the learning and understanding of figurative compared to literal language. However, conventional metaphors (e.g., *bright student*) have been found to require much less processing resources than novel metaphors (Bambini, Gentili, Ricciardi, Bertinetto, & Pietrini, 2011; Cardillo, Watson, Schmidt, Kranjec, & Chatterjee, 2012; Schmidt & Seger, 2009). Conventional metaphorical language is processed as quickly as literal language, and these conventional metaphors are easily understood and integrated into our everyday language (Giora, 1999; Glucksberg, 1998; Pynte, Besson, Robichon, & Poli, 1996).

Whilst evidence supports a role of sleep in the processing and consolidation of emotional information, no studies to date have investigated whether sleep supports the learning of and memory for novel emotional stimuli. Previous studies of the effect of emotion and sleep on memory have focused on stimuli that already have emotional content for the participant (e.g., Chambers & Payne, 2014; Nishida et al., 2009; Payne et al., 2008). Furthermore, it is not yet known when and how the emotional content of figurative language comes into being – whether this is immediate or requires a period of consolidation. The current study therefore investigates whether sleep differentially affects memory for conventional and novel metaphorical word pairs of negative, positive, and neutral valence. We hypothesised that there would be an overall benefit of sleep compared to wake on both memory for novel metaphorical definitions, and accurate recognition of to-be-remembered word pairs. Furthermore, evidence suggests that difficult tasks benefit greatly from the consolidation effects of sleep (Sio, Monaghan, & Ormerod, 2013; Stickgold & Walker, 2004). Since previous findings indicate that conventional metaphors are easily understood and require less processing resources than novel ones (e.g., Cardillo et al., 2012; Bambini et al., 2011; Schmidt &

Seger, 2009), we predict that sleep will particularly enhance memory for novel metaphorical word pairs to a greater extent than conventional metaphors, leading to a greater difference in performance between the sleep and wake groups specifically for novel metaphors. Similarly, novel metaphorical word pairs differed in their sensibility, with some definitions matching the novel word pair more congruently than others. Thus, we hypothesise that those word pairs with incongruent definitions will be more difficult to learn and remember, and so we expect to see a greater benefit of sleep than wake for novel incongruent metaphorical word pairs.

Since evidence indicates increased memory for emotional stimuli (Kensinger & Corkin, 2003; Adelmann & Estes, 2013), we also predict that emotionality of the word pairs would lead to both increased accurate recognition of previously seen word pairs, as well as increased learning of the definitions of novel metaphorical word pairs. Due to inconsistencies within the literature as to the effects of valence versus arousal, we may expect to see an increase in memory performance for word pairs rated high in negative valence and arousal (Kensinger & Corkin, 2003), or increases in memory for word pairs of both negative and positive valence, independent of arousal (Adelmann & Estes, 2013).

Finally, if sleep affects emotional memory processing (van der Helm & Walker, 2009; Walker, 2009), then we would expect sleep to enhance both accurate recognition of word pairs and increase learning of definitions for those word pairs rated as more emotional. Furthermore, there are suggestions that not only does sleep, and in particular REM sleep, aid in emotional memory consolidation, but it also changes the affective tone of the to-be-remembered information. The sleep to forget, sleep to remember (SFSR) hypothesis (van der Helm & Walker, 2009) suggests that emotional memories do persist over time, but the emotional response to such information is reduced as a consequence of sleep. Research conducted by van der Helm, Yao, Dutt, Rao, Saletin, & Walker (2011) supports the SFSR hypothesis; REM sleep in particular was associated with decreased emotional responses to pictures. In the present study, we therefore not only predict an increase in memory for emotional word pairs after sleep, but also expect ratings of arousal and valence of the word pairs to be reduced after sleep, whereas the wake group will show no difference in ratings over time.

2. Methods

2.1 Participants

Forty-nine Lancaster University students took part in the study for course credit or payment of £7. All participants gave informed consent and were fully debriefed at the end of the study. Non-native British English speakers (*n* = 6) were removed from the analysis, as well as one participant who self-reported being in bed for less than 6 hours (cut-off based on previous studies of sleep and memory, e.g., Diekelmann, Landolt, Lahl, Born, & Wagner, 2008). Two participants were removed due to computer error during completion of the experiment. This led to the inclusion of 40 participants (34 female) with a mean age of 20.8 years (SD = 3.01, range = 18 to 32).

2.2 Materials

We collated 71 conventional and 67 novel metaphorical word pairs from a range of sources; 56 conventional and 59 novel word pairs were taken from Liu, Connell and Lynott (under review), two conventional word pairs from Mashal & Itkes (2012), and five conventional and four novel word pairs taken from Forgács, Lukás, and Pléh (2014). Eight conventional metaphors and four novel metaphors were also added.

Using these word pairs, we collected ratings on a 7-point scale measuring sensibility (the extent to which a word pair made sense), metaphoricity (the extent to which a word pair expressed a non-literal expression), familiarity (extent of exposure with the word pair), emotional valence (from negative to positive), and arousal (from non-arousing to arousing). We collected ratings for each word pair from 15 participants who did not take part in the main experiment, and 13 of those same participants completed a second questionnaire providing a definition of what they thought the meaning of each novel word pair was.

Based on these questionnaires, we extracted 15 conventional metaphorical word pairs and 30 novel metaphorical word pairs. For conventional word pairs, we chose those word pairs that were rated as high in metaphoricity, sensibility, and familiarity. To ensure that novel word pairs were not previously known by participants, we chose novel word pairs rated low in familiarity. For both conventional and novel word pairs, we ensured that a third of the word pairs were rated as high in negative valence, a third high in positive valence, and a third rated as neutral. We chose negative and positive word pairs that had similar ratings of arousal, whereas neutral word pairs were rated as lower in arousal. Thus, each of the three metaphor conditions (conventional, novel congruent, novel incongruent) were split by valence, so that there were five word pairs of negative valence, five word pairs of positive valence, and five of neutral valence in each condition.

The novel word pairs were then split into two categories; 15 word pairs in which we assigned a meaning that was congruent (the meaning assigned to the word pair was similar to the meanings provided by participants in the questionnaire, and so made sense; *e.g., "jingling satisfaction"* was paired with the meaning *"feeling very pleased and proud of yourself"*), and 15 word pairs that were assigned a meaning that was incongruent (the word pair was given a meaning that differed from those provided by participants in the questionnaire, and so did not make sense; *e.g., "jingling satisfaction"* was paired with *"a generous reward"*). Whether a word pair was in the congruent or incongruent condition was counterbalanced, so that half of participants saw one set of word pairs with a congruent meaning, and the other half saw the same set of word pairs with an incongruent meaning. Independent samples t-tests on positive lists showed no significant differences between the two groups of novel word pairs for sensibility $t(8) = .103$, SE = .658, $p = .920$, metaphoricity $t(8) = -.110$, SE = .474, $p = .915$, valence $t(8) = .451$, SE = .266, $p = .664$, arousal $t(8) = -.253$, SE = .531, $p = .807$, or familiarity $t(8) = .225$, SE = .532, $p = .827$. There were no significant differences between neutral word pairs for sensibility $t(8) = -.433$, SE = .748, $p = .676$, metaphoricity $t(8) = 1.003$, SE = .399, $p = .345$, valence $t(8) = -1.134$, SE = .270, $p = .290$, arousal $t(8) = -.615$, SE = .492, $p = .556$, or familiarity $t(8) = -.378$, SE = .492, $p = .715$. Finally, there were also no significant differences between the two groups of negative word pairs for sensibility $t(8) = -.719$, SE = .634, $p = .492$, metaphoricity $t(8) = .412$, SE = .262, $p = .691$, valence $t(8) = 1.722$, SE = .310, $p = .123$, arousal $t(8) = -2.239$, SE = .649, $p = .056$, or familiarity $t(8) = -.973$, SE = .450, $p = .359$. See Table 1 for mean ratings of sensibility, metaphoricity, valence, arousal, and familiarity across groups of novel word pairs.

Table 6.1

Means and standard deviation of the two lists of novel word pairs, split by valence, for ratings of sensibility, metaphoricity, valence, arousal, and familiarity.

	Positive		Negative		Neutral	
	List 1	List 2	List 1	List 2	List 1	List 2
Sensibility	4.080(.863)	4.012(1.192)	4.038(1.132)	4.494(.853)	3.146(1.010)	3.470(1.334)
Metaphoricity	4.600(.789)	4.652(.708)	4.813(.360)	4.706(.462)	4.466(.739)	4.066(.499)
Valence	1.160(.304)	1.280(.511)	-1.387(.479)	-1.920(.500)	-.400(.346)	-.093(.496)
Arousal	3.464(.758)	3.598(.913)	2.838(.988)	4.292(1.064)	1.880(.387)	2.108(.734)
Familiarity	2.094(.996)	1.974(.651)	1.948(.715)	2.386(.708)	1.814(.768)	2.000(.789)

We also conducted independent samples t-tests to assess the differences in ratings of sensibility, metaphoricity, valence, arousal, and familiarity between conventional and novel word pairs between positive, negative, and neutral word pairs (see Table 2 for means and standard deviation of each measure). The results revealed a significant difference in sensibility between conventional and novel metaphors for positive word pairs $t(13) = 5.722$, SE = .446, $p < .001$, negative word pairs $t(8) = 4.217$, SE = .457, $p = .001$, and neutral word pairs $t(13) = 5.744$, SE = .528, $p < .001$. Conventional word pairs were rated as making more sense. Conventional word pairs were also rated as significantly more familiar than novel word pairs for positive $t(13) = 8.264$, SE = .442, $p < .001$, negative $t(13) = 7.875$, SE = .374, $p < .001$, and neutral metaphors $t(13) = 7.254$, SE = .507, $p < .001$.

For positive word pairs, there was also a significant difference in metaphoricity between conventional and novel metaphors $t(13) = 2.793$, SE = ..330, $p = .015$, with novel metaphors rated as lower in metaphoricity. There was no difference between conventional and novel word pairs for neutral $t(13) = 1.360$, SE = .355, $p = .197$, or negative metaphors $t(13) = .553$, SE = .322, $p = .590$. The results also revealed a significant difference in ratings of valence between positive conventional and novel metaphors *t*(13) = 3.653, SE = .197, *p* = .003. Participants rated the novel word pairs as more neutral in valence than the conventional word pairs. There was no difference in ratings of valence for neutral *t*(13) = 1.551, SE = .279, *p* = .145 or negative word pairs *t*(13) = .022, SE = .304, *p* = .983. Since metaphors are often highly emotive (Fainsilber & Ortony, 1987) it could be that since the positive novel word pairs were rated as lower in metaphoricity, this explains the lower valence rating also evident for the novel word pairs.

Ratings of arousal did not significantly differ between conventional and novel metaphors for positive *t*(13) = .637, SE = .457, *p* = .535, neutral *t*(13) = 1.432, SE = .358, *p* = .176, or negative word pairs *t*(13) = .823, SE = .650, *p* = .425.

Table 6.2

Means and standard deviation of conventional and novel word pairs, split by valence, for ratings of sensibility, metaphoricity, valence, arousal, and familiarity.

	Positive		Negative		Neutral	
	Conventional	Novel	Conventional	Novel	Conventional	Novel
Sensibility	6.572(.248)	4.020(.965)	6.252(.376)	4.326(.970)	6.342(.396)	3.308(1.128)
Metaphoricity	5.428(.443)	4.506(.661)	4.838(.856)	4.660(.417)	4.748(.683)	4.266(.631)
Valence	2.053(.288)	1.333(.387)	-1.733(.678)	-1.740(.491)	.187(.649)	-.247(.434)
Arousal	3.802(.944)	3.511(.780)	4.320(1.442)	3.785(1.054)	2.506(.815)	1.994(.566)
Familiarity	5.640(.880)	1.987(.772)	5.146(.613)	2.200(.712)	5.586(1.246)	1.907(.741)

We also ensured that positive, negative, and neutral word pairs significantly differed in overall valence ratings, for both conventional and novel metaphors. For conventional metaphors, negative word pairs were rated as significantly more negative than both positive *t*(8) = 11.489, SE = .330, *p* < .001, and neutral word pairs *t*(8) = -4.574, SE = .420, *p* = .002. Positive word pairs were rated as significantly more positive than neutral word pairs *t*(8) = 5.879, SE = .317, *p* < .001. Similarly, for novel metaphors, positive word pairs were rated as significantly more positive than neutral word pairs *t*(18) = 8.584, SE = .184, *p* < .001, and negative word pairs significantly differed from both positive *t*(18) = 15.531, SE = .198, *p* < .001, and neutral word pairs *t*(18) = -7.200, SE = .207, *p* < .001.

We also analysed whether positive, negative, and neutral word pairs differed in ratings of arousal. For both conventional and novel metaphors, positive word pairs were rated as significantly higher in arousal than neutral word pairs *t*(8) = 2.323, SE = .558, *p* = .049; *t*(18) = 4.979, SE = .305, *p* < .001. Negative word pairs were also rated as higher in arousal than neutral word pairs for both conventional *t*(8) = 2.448, SE = .741, *p* = .040, and novel metpahors *t*(18) = 4.733, SE = .378, *p* < .001. Negative and positive word pairs did not significantly differ in ratings of arousal for conventional *t*(8) = -.672, SE = .771, *p* = .521, or novel metaphors *t*(18) = -.661, SE = .415, *p* = .517.

During testing 90 word pairs were presented to participants; the 45 word pairs that they were exposed to in the first session (previously seen), plus 45 new word pairs that were either conventional metaphors (15), or novel metaphors (30) (previously unseen). Unseen word pairs were matched to seen word pairs in terms of valence, resulting in 15 unseen positive word pairs, 15 unseen negative word pairs, and 15 unseen neutral word pairs (see Table 3 for mean ratings of sensibility, metaphoricity, valence, arousal, and familiarity for unseen word pairs). For positive, neutral, and negative conventional word pairs, seen and unseen metaphors did not differ in sensibility *t*(8) = .587, SE = .157, *p* = .574; *t*(8) = .035, SE = .286, *p* = .973; *t*(8) = -

.733, SE = .221, *p* = .484, metaphoricity *t*(8) = 1.187, SE = .315, *p* = .269; *t*(8) = .137, SE = .495, *p* = .894; *t*(8) = .052, SE = .466, *p* = .960, valence *t*(8) = .507, SE = .263, *p* = .626; *t*(8) = .469, SE = .341, *p* = .652; *t*(8) = .454, SE = .382, *p* = .662, arousal *t*(8) = -.146, SE = .533, *p* = .887; *t*(8) = -1.080, SE = .728, *p* = .312; *t*(8) = .032, SE = .802, *p* = .975, or familiarity *t*(8) = -.025, SE = .474, *p* = .980; *t*(8) = .441, SE = .667, *p* = .671; *t*(8) = -2.041, SE = .320, *p* = .076.

For positive, neutral, and negative novel word pairs, seen and unseen metaphors did not differ in sensibility *t*(18) = .669, SE = .400, *p* = .512; *t*(18) = -.578, SE = .492, *p* = .571; *t*(18) = -.269, SE = .543, *p* = .791. For ratings of metaphoricity, positive and neutral word pairs did not differ *t*(18) = -.996, SE = .261, *p* = .332; *t*(18) = -.638, SE = .252, *p* = .532, however negative word pairs did *t*(18) = -2.320, SE = .212, *p* = .032. Unseen word pairs were rated higher in metaphoricity. Positive word pairs differed in valence *t*(18) = 2.244, SE = .152, *p* = .038; seen word pairs were rated as higher in valence than unseen word pairs. Neutral and negative word pairs did not differ in valence *t*(18) = -.722, SE = .203, *p* = .479; *t*(18) = -1.345, SE = .203, *p* = .195. Positive, neutral, and negative novel metaphors did not differ in arousal *t*(18) = -.063, SE = .364, *p* = .950; *t*(18) = -1.245, SE = .353, *p* = .229; *t*(18) = -.428, SE = .470, *p* = .674, or familiarity *t*(18) = .065, SE = .309, *p* = .949; *t*(18) = -.794, SE = .412, *p* = .437; *t*(18) = -1.648, SE = .393, *p* = .117.

Table 6.3

Means and standard deviation of unseen conventional and novel word pairs, split by valence, for ratings of sensibility, metaphoricity, valence, arousal, and familiarity.

	Positive		Negative		Neutral	
	Conventional	Novel	Conventional	Novel	Conventional	Novel
Sensibility	6.480(.248)	3.752(.820)	6.414(.320)	4.472(1.418)	6.332(.503)	3.592(1.069)
Metaphoricity	5.054(.548)	4.766(.494)	4.814(.593)	5.152(.526)	4.680(4.427)	.872(.490)
Valence	1.920(.513)	.993(.282)	-1.907(.518)	-1.467(.415)	.027(.402)	-.100(.473)
Arousal	3.880(.729)	3.534(.848)	4.294(1.068)	3.986(1.048)	3.292(1.408)	2.434(.963)
Familiarity	5.652(.590)	1.967(.599)	5.800(.371)	2.847(1.017)	5.292(.819)	2.234(1.071)

Participants in the sleep condition were required to wear an Actigraph sleep monitor overnight to measure time spent asleep, sleep quality, and number of awakenings during the night. Participants were also asked to complete a questionnaire on their general sleep habits and caffeine and alcohol intake, as well as completing the Self-Assessment Manikin (SAM; Bradley & Lang, 1994), and the Stanford Sleepiness Scale (SSS; Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973), at the start of each session. The SAM mood questionnaire was given to assess participants' ratings of valence (negative-positive), activation (calm-excited), and control (dominated-dominant) on a scale from 1-9. The SSS was given to assess participants' self-reported levels of sleepiness at both time periods, on a scale from 1-7.

2.3 Procedure

Participants were randomly allocated to the sleep or wake condition, with 20 participants in each condition. Those in the wake group were exposed to the word pairs and definitions at 9am and then tested the same day at 9pm after a day of wakefulness. Participants in the sleep group were exposed to word pairs and definitions at 9pm, and then tested at 9am after a night of sleep.

During the first session, participants first completed the sleep habits and alcohol and caffeine intake questionnaire, the SSS, and the SAM mood scale. Participants were then seated approximately 60cm from the computer screen and were exposed to the 15 conventional, 15 novel congruent, and 15 novel incongruent word pairs and definitions. Participants were presented with each word pair and definition one at a time for 5s each. After 5s, the definition would then disappear from the screen, and participants were required to first provide a rating of valence and arousal for the word pair. Valence was rated on a scale from negative (-3) to positive (+3), with zero meaning completely neutral, and arousal was rated on a scale from not at all arousing (1) to completely arousing (7). Participants were asked to press the corresponding key on the keyboard to give their response. Participants were then required to reproduce the definition of the word pair by typing the meaning into the computer using the keyboard. Participants were not given a time limit to complete the ratings or definitions. Each word pair and definition was presented in random order and was presented once only.

After all 45 word pairs and definitions had been presented to participants, we again gave participants each word pair one at a time in random order, and asked them to type the definition of each word pair again. Participants were instructed not to guess, and to type “I don’t know” if they could not remember the definition. Each word pair remained on the screen until participants entered their definition, and participants were not given a time limit to complete these definitions.

Participants then left the lab, and those in the wake condition were instructed not to nap throughout the day. Twelve hours later, after a period of overnight sleep or daytime wake, participants returned for the testing phase of the experiment. In this second session, participants were again asked to complete the sleep habits and alcohol and caffeine questionnaire, the SSS and the SAM mood scale. Participants were then presented with a central fixation cross for 500ms, before one of the word pairs was presented to the centre of the screen. Participants were asked to decide whether they recognised the word pair as one that was presented to them in the first session *(old)* or not presented in the first session *(new)*. However, participants made their decision on whether the word pair was old or new on a scale from one to six based on their confidence, with the press of 1 indicating definitely old, and the press of 6 indicating definitely new. The word pair remained on the screen until participants made a decision. Participants were not given a response deadline, but were asked to respond as quickly and accurately as possible. Once participants had made their decision, they were then presented with another fixation point for 500ms, before another word pair appeared on the screen. The presentation order of each word pair was randomised for each participant.

After participants had seen all word pairs, they were then given each of the previously seen word pairs again, one at a time. Participants were first asked to rate the valence and arousal of the word pair using the same rating scales as in the first session, and were then asked to provide the definition of the word pair by typing into the box provided for them on the computer screen using the keyboard. Participants were instructed not to guess, and to type “I don’t know” if they did not know the answer. The word pair remained on the screen until participants gave the ratings and definition, and participants were not given a response deadline.

2.4 Data Analysis

Data were analysed using the package lme4 (Bates, Maechler, & Bolker, 2012) in R (R Core Team, 2015). For binary outcomes (recall and recognition accuracy), general linear mixed-effects models were used. For confidence ratings and response time (RT) data, linear mixed-effects analysis was performed. For all analyses, to determine the best-fitting model fixed effects were added incrementally. The model was initially fitted with random effects of participants and items (word pairs), and then fixed effects were added to the model one at a time. Each model was compared to the previous model using likelihood ratio tests, and fixed effects remained in the final model if they contributed significantly to model fit.

For recall of definitions, whether the definition provided by participants was correct or incorrect was determined by two scorers independently. Responses were scored as correct if the definition was the same or similar to that learned during training. If synonyms of the definition were used, the definition was scored as correct. Where the scorers differed in their decision, the participant's response was discussed between scorers, and a final decision was made. Both recall and recognition accuracy were coded such that a correct response was given a score of 1, and an incorrect response was given a score of 0.

We asked participants to rate how confident they were in their responses to word pairs, with confidence ratings ranging from 1 (confident that the word pair was previously seen) to 6 (confident that the word pair was new).To analyse differences in confidence ratings, we first reversed participants ratings of unseen word pairs, so that a score of 1 would show high confidence in accurately accepting seen word pairs, but also accurately rejected unseen word pairs, and a score of 3 would show low confidence. This allowed us to more easily compare confidence ratings of seen and unseen word pairs.

RT's for correct responses in the recognition task were analysed. Response times exceeding 2.5 standard deviations from the mean were removed from the analysis, leading to the removal of 2.439% of the data, and leaving 3157 data points. Based on test of normality, we reduced skew in the RT distribution by transforming latencies to $\log_{10}$(RT), and analysed the logRTs, based on suggestions from Baayen, Feldman, and Schreuder (2006).

To test performance between novel versus conventional word pairs, and between novel congruent and novel incongruent word pairs, we coded the "type of metaphor" categorical variable using Helmert contrasts. For novel versus conventional metaphors (contrast A), conventional word pairs = 1, and both novel congruent and novel incongruent word pairs = -.5. To compare novel congruent versus novel incongruent word pairs (contrast B), conventional word pairs = 0, novel congruent word pairs = -.5, and novel incongruent word pairs = .5.

We also assessed multicollinearity of emotion variables to investigate whether these predictors are highly correlated with each other. First, we dummy coded the emotionality factor, such that for negative word pairs, negative stimuli had a score of 1 and positive and neutral stimuli had a score of 0. For positive word pairs, positive stimuli had a score of 1, and for neutral stimuli neutral word pairs had a score of 1. We then calculated the variance inflation factor (VIF) for each level of emotionality, valence at training and testing, and arousal at training and testing. All VIF scores were below 5, which indicates a low level of collinearity (see Table 4 for VIF scores).

Table 6.4.

Variance Inflation Factors for all emotion variables, providing a measure of multicollinearity.

	Variance Inflation Factor
Negative	3.854
Positive	2.156
Neutral	1.935
Valence at training	3.089
Valence at testing	2.574
Arousal at training	1.416
Arousal at testing	1.578

3. Results

3.1 Self-Assessment Manikin and Sleepiness Scale

Participants filled out a SAM mood questionnaire and Stanford Sleepiness Scale prior to taking part in both the training and testing session. Mean scores of sleepiness, valence, activation, and control at both time points are reported in Table 5.

Table 6.5

Means and standard deviation of sleepiness score and SAM mood score prior to both sessions.

	Before training		Before testing	
	Sleep Group	Wake Group	Sleep Group	Wake Group
Mean(SD) sleepiness score	3.556(1.117)	2.773(1.445)	2.944(1.079)	3.273(1.355)
Mean(SD) valence score	6.111(1.197)	6.409(1.231)	6.167(1.385)	6.409(1.851)
Mean(SD) activation score	3.556(1.301)	3.545(1.196)	3.500(1.303)	4.545(1.672)
Mean(SD) control score	4.944(1.581)	5.591(1.338)	5.389(1.639)	5.455(1.530)

To test whether mood and sleepiness had no effect on overall recall of definitions at testing, we performed General Linear Mixed-Effects Models with recall memory accuracy at testing as the dependent variable. Adding SAM valence score did not improve model fit compared to a model with only random effects at either time 1 ($x^2(1) = .987, p = .321$) or time 2 ($x^2(1) = .163, p = .687$). Activation score before both training (x^2 $(1) = .006, p = .941$) and testing ($x^2(1) = 1.125, p = .289$) did not improve model fit, nor did control, ($x^2(1) = .704, p = .402$; $x^2(1) = 1.270, p = .260$). Similarly, sleepiness ratings before both training and testing did not significantly affect performance accuracy ($x^2(1) = .702, p = .402$; $x^2(1) = 1.945, p = .163$). This suggests no overall significant difference in performance accuracy dependent on ratings of mood or sleepiness.

We then wanted to test whether ratings of mood and sleepiness affected performance differently in the sleep versus wake group. Adding the interaction term group (sleep/wake) by valence score did not improve model fit at either time 1 ($x^2(3) = 1.743, p = .628$) or time 2 ($x^2(3) = 1.063, p = .786$), nor did group by activation score ($x^2(3) = 2.469, p = .481$; $x^2(3) = .2.283, p = .516$), or group by control score ($x^2(3) = 1.874, p = .599$; $x^2(3) = 2.661, p = .447$). Ratings of sleepiness before both training and testing did not significantly affect performance accuracy in the sleep versus wake group ($x^2(3) = 1.625, p = .654$; $x^2(3) = 2.904, p = .407$). Thus, participants' mood and sleepiness ratings in the wake versus sleep group did not significantly affect performance accuracy.

We also tested whether SAM mood score or sleepiness score influenced overall performance accuracy in the recognition task (accurate recognition of previously seen word pairs and accurate rejection of unseen word pairs). Adding SAM valence score did not improve model fit compared to a model with only random effects at either time 1 ($x^2(1) = 1.067$, $p = .302$) or time 2 ($x^2(1) = .342$, $p = .559$). Activation score before both training ($x^2(1) = .037$, $p = .848$) and testing ($x^2(1) = .467$, $p = .495$) did not improve model fit, nor did control, ($x^2(1) = .387$, $p = .534$; $x^2(1) = 3.106$, $p = .078$). Similarly, sleepiness ratings before both training and testing did not significantly affect performance accuracy ($x^2(1) = .534$, $p = .465$; $x^2(1) = 2.162$, $p = .142$). This suggests no overall significant difference in performance accuracy dependent on ratings of mood or ratings of sleepiness.

Finally, we tested whether ratings of mood and sleepiness affected recognition performance differently in the sleep versus wake groups. Since the main effect of group (sleep/wake) was found to be significant, we compared the interaction term group (sleep/wake) by SAM score/sleepiness score to a model including only the main effect of group. Adding the interaction term group by valence score did not improve model fit at either time 1 ($x^2(2) = 3.794$, $p = .150$) or time 2 ($x^2(2) = .338$, $p = .845$), nor did group by activation score ($x^2(2) = 1.162$, $p = .447$; $x^2(2) = 3.366$, $p = .186$). However, although group by control score at time 1 did not improve model fit, ($x^2(2) = 2.326$, $p = .313$), at time 2 we found a significant interaction between group and control score ($x^2(2) = 9.688$, $p = .008$). In the wake group, increased ratings of control led to increased accuracy (estimate = .270, SE = .082, z = 3.282, $p = .001$), whereas ratings of control did not influence accuracy in the sleep group (estimate = .002, SE = .080, z = .022, $p = .983$). We therefore included control score at time 2 as a random effect within our main analyses of recognition accuracy. Ratings of sleepiness both before training and testing did not significantly affect performance accuracy in the sleep versus wake group ($x^2(2) = 5.343$, $p = .069$; $x^2(2) = 1.862$, $p = .394$).

3.2 Recall of Definitions

We analysed participant's recall of definitions at two time points; immediately after exposure to the word pair and definition in the training session (immediate recall), and in the testing session after a period of sleep or wake (delayed recall). First, we analysed recall accuracy immediately after exposure, to test whether participants were paying attention to the word pairs and definitions and were able to recall definitions accurately in both the sleep and wake group. We then analysed delayed recall performance accuracy, first to investigate effects of metaphor type, arousal, and valence on recall performance, and then to investigate the specific effects of sleep versus wake on metaphor learning and emotionality.

3.2.1 Immediate Recall (at training)

During training participants were first presented with a word pair and its definition and were asked to reproduce the definition immediately afterwards. This was done to ensure that participants were paying attention to the stimuli and were able to accurately reproduce the definitions. We therefore wanted to ensure that there was no difference in how well participants could remember the definitions at this point between the sleep and wake group. We ran a general linear mixed-effects model, with recall accuracy at time 1 as the dependent variable. Adding the fixed factor of group to a model with only random effects did not improve model fit ($x^2(1) = 1.202$, $p = .273$), indicating no significant difference between the sleep and wake group. Indeed, mean proportion correct in the sleep and wake group was 0.970 (SD = .170) and 0.954 (SD = .211) respectively, demonstrating high recall performance at time 1 in both groups.

3.2.2 Delayed Recall (at test)

We performed a general linear mixed-effects model analysis, modelling the probability of recall accuracy of definitions. We included random effects of participants and items (word pairs), as well as fixed effects of type of metaphor (conventional, novel congruent, novel incongruent), emotionality (negative, positive, neutral), participants' valence and arousal score for each word pair at training and testing, group (sleep versus wake), participants' recall performance accuracy at time 1, as well as all interaction terms.

First, we investigated how well participants were able to remember definitions depending on the type of metaphor. Adding the main effect of type of metaphor to a model with only random effects significantly improved model fit, ($x^2(2) = 503.59$, $p = 2.2 \times 10^{16}$). As expected, the definitions of conventional metaphorical word pairs were recalled significantly more accurately than novel congruent word pairs, and definitions of novel incongruent word pairs were recalled least accurately.

We then added main effects of emotionality (negative, positive, and neutral) and ratings of valence and arousal to the model, to investigate whether any emotionality effects influenced recall performance. We found that participants' ratings of arousal at testing significantly improved model fit, ($x^2(2) = 15.235$, $p = 9.495 \times 10^5$). The results indicated a linear relationship, with increased ratings of arousal leading to greater performance accuracy (estimate = 1.044, SE = .321, z = 3.254, p = .001). There was also a main effect of recall performance at time 1 ($x^2(2) = 62.686$, $p = 2443 \times 10^{14}$); with increased performance at time 1 leading to increased performance accuracy at test (estimate = 1.007, SE = .328, z = 3.073, p = .002).

No other main effects were found to be significant. Adding interaction terms with sleep versus wake group did not significantly improve model fit, suggesting no specific effects of sleep on learning of metaphorical word pairs, or on emotional memory performance. The final model is reported in Table 6.

Table 6.6

Summary of general linear mixed effects model of accurate recall of definitions, including fixed effects of type of word pair, arousal rating score at testing, and accurate recall of definitions at training.

	Estimated		Wald confidence intervals			
Fixed effects	coefficient	SE	2.50%	97.50%	z	Pr(>|z|)
Type (novel congruent vs. conventional)	-0.872	0.348	-1.554	-0.190	-2.506	0.012
Type (novel incongruent vs. conventional)	-4.625	0.396	-5.402	-3.848	-11.669	<.001
Arousal at testing	1.007	0.328	0.365	1.650	3.073	0.002
Definition accuracy at training	2.512	0.473	1.584	3.439	5.307	<.001
Random effects	Name	Variance	Std. Dev			
Subject effect on intercepts	(Intercept)	0.914	0.956			
Item effect on intercepts	(Intercept)	0.977	0.988			
	AIC	BIC	logLik	deviance		
	1537.5	1603.4	-756.8	1513.5		

1800 observations, 40 participants, 45 items

3.3 Recognition of Word Pairs

We measured participant’s performance accuracy, in terms of their ability to accurately recognise word pairs exposed to during training, and accurately reject previously unseen word pairs. We also analysed participants’ confidence in their responses, rated from 1-6, with a rating of one corresponding to highest confidence that a word pair was previously seen, and a rating of 6 corresponding to highest confidence that a word pair was unseen. Thus, ratings of three and four corresponded to the lowest confidence ratings. Finally, we also analysed response time (RT) for accurate responses.

For each dependent variable, we first ran mixed effects models including only responses to previously seen word pairs, to investigate differences between novel congruent and novel incongruent word pairs. We then investigated seen and unseen word pairs, to examine overall differences in performance between novel and conventional metaphorical word pairs, using Helmert contrasts to analyse differences between conventional and novel word pairs (contrast A), and novel congruent versus novel incongruent word pairs (contrast B).

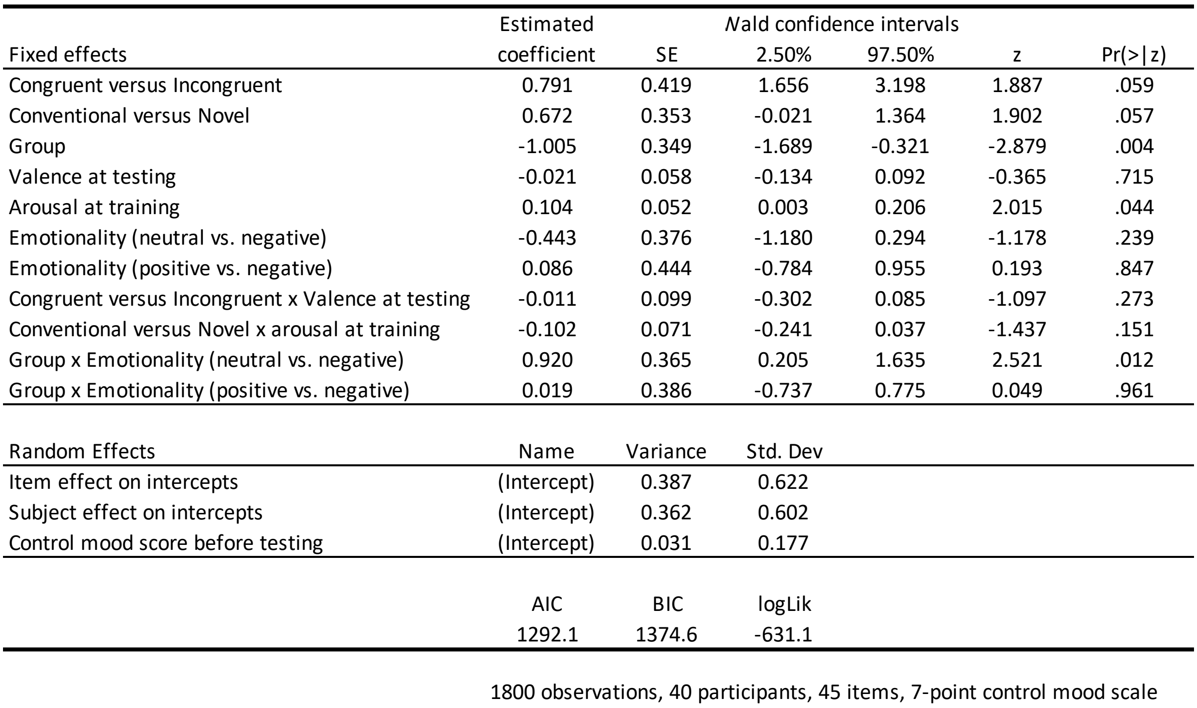
We took a hypothesis-driven approach to model building. If sleep aids in memory for the more difficult to remember information, we would expect to see a greater effect of sleep versus wake for novel incongruent word pairs than novel congruent, and for novel than conventional word pairs. We also included participant ratings of valence and arousal at training and at testing to the model. If sleep affects memory for emotional information, then we would expect to see an interaction between arousal and sleep versus wake, whereas if there are differential effects of negative and positive emotion, then we may expect an interaction between valence and sleep versus wake group.

3.3.1 Accuracy

We performed General Linear Mixed-effects Models, modelling the probability of response accuracy for previously seen word pairs. We included random effects of participants and items (word pairs), as well as the random effect of SAM control score at testing. We added main effects of type of word pair coded as Helmert contrasts (first comparing conventional to novel word pairs, and then congruent versus incongruent word pairs), group (sleep, wake), emotionality (negative, positive, neutral), and valence and arousal score at both training and testing, as well as interaction terms using a hypothesis-driven approach to model building. We then conducted exploratory analyses of all other interaction effects. See Table 7 for the final model summary, and Figure 1 for means and standard error across conditions.

Table 6.7

Summary of general linear mixed effects model of recognition accuracy of previously seen word pairs.



	Estimated		Wald confidence intervals			
Fixed effects	coefficient	SE	2.50%	97.50%	z	Pr(>|z)
Congruent versus Incongruent	0.791	0.419	1.656	3.198	1.887	.059
Conventional versus Novel	0.672	0.353	-0.021	1.364	1.902	.057
Group	-1.005	0.349	-1.689	-0.321	-2.879	.004
Valence at testing	-0.021	0.058	-0.134	0.092	-0.365	.715
Arousal at training	0.104	0.052	0.003	0.206	2.015	.044
Emotionality (neutral vs. negative)	-0.443	0.376	-1.180	0.294	-1.178	.239
Emotionality (positive vs. negative)	0.086	0.444	-0.784	0.955	0.193	.847
Congruent versus Incongruent x Valence at testing	-0.011	0.099	-0.302	0.085	-1.097	.273
Conventional versus Novel x arousal at training	-0.102	0.071	-0.241	0.037	-1.437	.151
Group x Emotionality (neutral vs. negative)	0.920	0.365	0.205	1.635	2.521	.012
Group x Emotionality (positive vs. negative)	0.019	0.386	-0.737	0.775	0.049	.961

Random Effects	Name	Variance	Std. Dev
Item effect on intercepts	(Intercept)	0.387	0.622
Subject effect on intercepts	(Intercept)	0.362	0.602
Control mood score before testing	(Intercept)	0.031	0.177

	AIC	BIC	logLik
	1292.1	1374.6	-631.1

1800 observations, 40 participants, 45 items, 7-point control mood scale

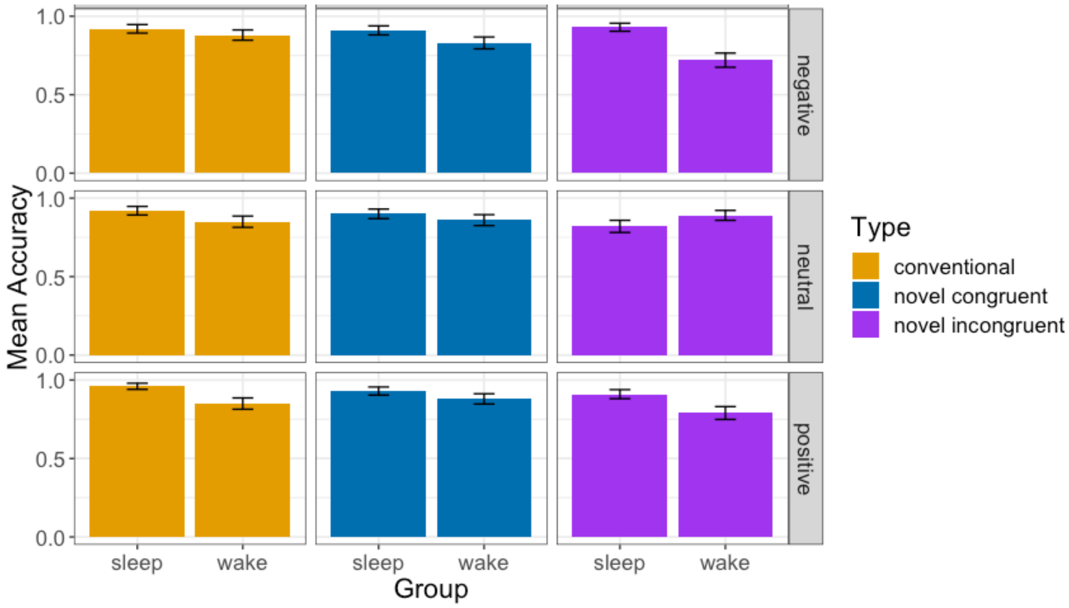


Figure 6.1. Mean accuracy for previously seen conventional, novel congruent, and novel incongruent word pairs in the sleep and wake group separately, split by emotionality. Error bars show ±1 SEM.

Previously Seen Word Pairs

Confirmatory Analysis

Adding the main effect of conventional versus novel word pairs (contrast A) to the model did not improve model fit ($x^2(1) = 1.771$, $p = .183$). Congruent versus incongruent word pairs (contrast B) did significantly improved model fit ($x^2(1) = 5.702$, $p = .026$). As predicted, novel congruent metaphors were accurately recognised significantly more than novel incongruent word pairs (estimate = .433, SE = .177, z = 2.444, $p = .015$).

Adding the main effect of group to the model also significantly improved model fit, ($x^2(1) = 4.927$, $p = .026$). The sleep group performed significantly more accurately than the wake group (estimate = .662, SE = .274, z = -2.414, $p = .016$). Arousal score at training ($x^2(6) = 10.106$, $p = .120$) and testing ($x^2(6) = 11.070$, $p =$

.086) did not improve model fit, nor did valence at training ($x^2(6) = 10.727$, $p = .097$) or testing ($x^2(6) = 10.252$, $p = .114$). Similarly, emotionality did not improve model fit ($x^2(8) = 13.479$, $p = .096$).

We then added interaction effects based on our hypotheses. We predicted an effect of sleep specifically for those word pairs that were more difficult to learn. However, adding the interaction between novel versus conventional word pairs and group did not improve model fit ($x^2(2) = 5.277$, $p = .071$), nor was the interaction between congruent versus incongruent word pairs and group ($x^2(1) = .127$, $p = .722$).

We then added the interaction term group by emotionality. We predicted that the sleep group would show increased performance accuracy for emotional word pairs. Adding the interaction to the model did improve model fit ($x^2(4) = 12.117$, $p = .017$). Thus, we first investigated the effect of emotionality on the sleep and wake group separately. For the sleep group, there was no significant difference in performance between neutral and negative word pairs (estimate = -.491, SE = .470, z = -1.046, p = .296), neutral and positive word pairs (estimate = -.846, SE = .497, z = -1.702, p = .089), or positive and negative word pairs (estimate = .354, SE = .510, z = .694, p = .488). For the wake group, there was no significant difference in performance between neutral and negative word pairs (estimate = .464, SE = .279, z = 1.664, $p = .096$), neutral and positive word pairs (estimate = .245, SE = .284, z = .861, $p = .389$), or positive and negative word pairs (estimate = .219, SE = .271, z = .810, $p = .418$). We then analysed the effect of group on negative, positive, and neutral word pairs. For negative word pairs, the wake group performed significantly less accurately than the sleep group (estimate = -1.056, SE = .328, z = -3.220, $p = .001$). Similarly, for positive word pairs, the wake group performed less accurately than the sleep group (estimate = -1.045, SE = .359, z = -2.906, $p = .004$). There was no significant difference between the sleep and wake group for negative word pairs (estimate = -.048, SE = .270, z = -.178, *p* = .859).

We also predicted that valence and arousal scores would interact with group. However, adding the interaction term group by valence at training (x^2(12) = 7.513, *p* = .131), arousal at training (x^2(12) = 14.789, *p* = .571), valence at testing (x^2(12) = 7.322, *p* = .292), and arousal at testing (x^2(12) = 18.250, *p* = .108) did not improve model fit.

Exploratory Analyses

We also analysed further interaction terms that were not part of our hypothesis. We added all possible 2-way and 3-way interactions one at a time to the model. Adding the interaction term novel versus conventional word pairs by arousal at training significantly improved model fit (x^2(13) = 34.766, *p* < .001). For novel word pairs, there was a significant effect of arousal at training (estimate = .155, SE = .057, z = 2.736, *p* = .006), accuracy increased as arousal rating increased. There was no effect of arousal for conventional word pairs (estimate = -.001, SE = .601, z = -.001, *p* = .999). Thus, the effect of emotionality on recognition accuracy is evident for newly learned metaphors, suggesting a potential immediate influence of emotion on learning and memory for novel information.

Seen versus Unseen Word Pairs

We then performed further General Linear Mixed-Effects Models, modelling the probability of response accuracy for both seen and unseen word pairs. We included random effects of participants and items (word pairs), as well as the random effect of SAM control score at testing. We added main effects of type of metaphor (novel, conventional), group (sleep, wake), emotionality (negative, positive, neutral), and seen versus unseen word pairs, plus all interaction terms. See Table 8 for the final model summary and Figure 2 for means and standard error.

Table 6.8

Summary of general linear mixed effects model of recognition accuracy of seen and unseen word pairs, including fixed effects of sleep versus wake group, seen versus unseen word pairs, emotionality, and the group by emotionality interaction.

	Estimated		Wald confidence intervals			
	coefficient	SE	2.50%	97.50%	z	Pr(>|z|)
Fixed effects	-0.076	0.262	-0.589	0.437	-0.289	.772
Group (wake vs. sleep)	0.561	0.174	0.219	0.903	3.217	.001
Seen versus Unseen word pairs	0.159	0.269	-0.368	0.685	0.590	.555
Emotionality (negative vs. neutral)	0.137	0.268	-0.389	0.663	0.510	.610
Emotionality (positive vs. neutral)	-0.678	0.282	-1.231	-0.125	-2.401	.016
Group x Emotionality (negative vs. neutral) interaction	-0.659	0.282	-1.212	-0.106	-2.336	.020
Group x Emotionality (positive vs. neutral) interaction						
Random effects	Name	Variance	Std. Dev			
Subject effect on intercepts	(Intercept)	0.245	0.495			
Item effect on intercepts	(Intercept)	0.366	0.605			
Control mood score before testing	(Intercept)	0.011	0.107			
	AIC	BIC	logLik			
	2260.9	2322.8	-1120.5	2240.9		

1800 observations, 40 participants, 45 items, 7-point control mood scale

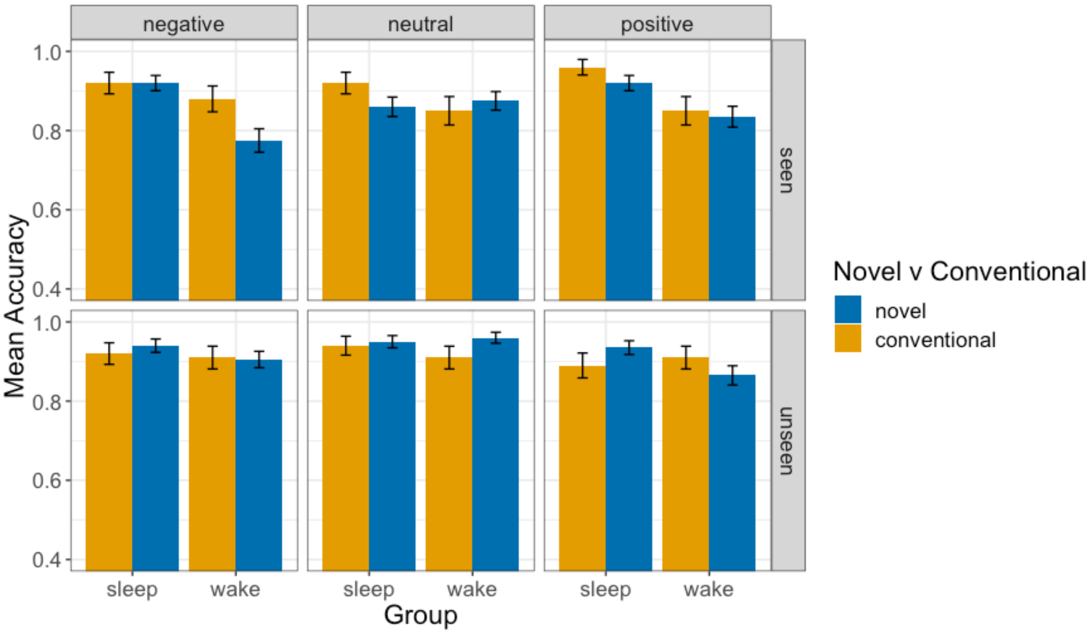


Figure 6.2. Mean accuracy for previously seen and unseen conventional and novel word pairs in the sleep and wake group separately, split by emotionality. Error bars show ±1 SEM.

Confirmatory Hypothesis

Adding the main effect of group to a model with only random effects significantly improved model fit ($x^2(1) = 6.365$, $p = .012$). The wake group performed significantly less accurately than the sleep group (estimate = -.536, SE = .200, z = -2.676, $p = .007$). Adding a main effect of seen versus unseen word pairs to the model also significantly improved model fit ($x^2(1) = 9.432$, $p = .002$), with unseen word pairs being accurately rejected more than seen word pairs were accurately accepted (estimate = .555, SE = .175, z = 3.175, $p = .002$). This suggests a possible response bias, with participants more likely to reject word pairs regardless of whether they were previously seen or unseen. Adding the main effect of novel versus conventional word pairs (contrast A) to the model did not improve model fit ($x^2(1) = .313$, $p = .576$), nor did the main effect of emotionality ($x^2(2) = 1.607$, $p = .448$).

We then added interaction terms based on our hypotheses. We predicted an interaction between group and type of word pair, such that the sleep group would show greater performance accuracy than the wake group specifically for more difficult to learn word pairs (the novel word pairs). Adding the interaction term group by novel versus conventional word pairs did not significantly improve model fit ($x^2(2) = .436$, $p = .804$).

The results also indicated a marginally significant interaction between group and emotionality ($x^2(4) = 8.625$, $p = .071$; see Figure 3). Based on our hypotheses, we predicted that the sleep group would show increased performance accuracy for emotional word pairs than the wake group. Thus, we analysed the difference in performance between the sleep and wake group for positive, negative, and neutral word pairs separately. For positive word pairs, the wake group performed less accurately than the sleep group (estimate = -.786, SE = .269, z = -2.918, $p = .004$). Similarly, for negative word pairs the wake group also performed with less accuracy than the sleep group, (estimate = -.776, SE = .256, z = -3.034, $p = .002$). For neutral word pairs, there was no significant difference between sleep and wake groups (estimate = -.106, SE = .218, z = -.489, $p = .625$). We further analysed the effect of emotionality on performance accuracy. We predicted that the sleep group would show increased performance accuracy for emotional than neutral word pairs. However, there was no significant effect of emotionality for either negative versus neutral stimuli (estimate = .174, SE = .305, z = .570, $p = .569$), positive versus neutral stimuli (estimate = .220, SE = .307, z = .718, $p = .472$), or positive versus negative stimuli (estimate = .047, SE = .312, z = .150, $p = .881$). We also investigated the effect of emotionality on performance accuracy in the wake group. In the wake group, there was no significant difference in recognition accuracy of positive versus negative word pairs (estimate = -.003, SE = .220, z = -.015, *p* = .988). However, both negative and positive word pairs were responded to less accurately than neutral word pairs (estimate = -.489, SE = .232, z = -2.105, *p* = .035; estimate = -.492, SE = .233, z = -2.116, *p* = .034). We then added all other 2-way, 3-way, and the 4-way interaction between group, seen/unseen, emotionality, and novel/conventional word pairs as exploratory analyses. No interactions significantly improved model fit (p > .05 for all).

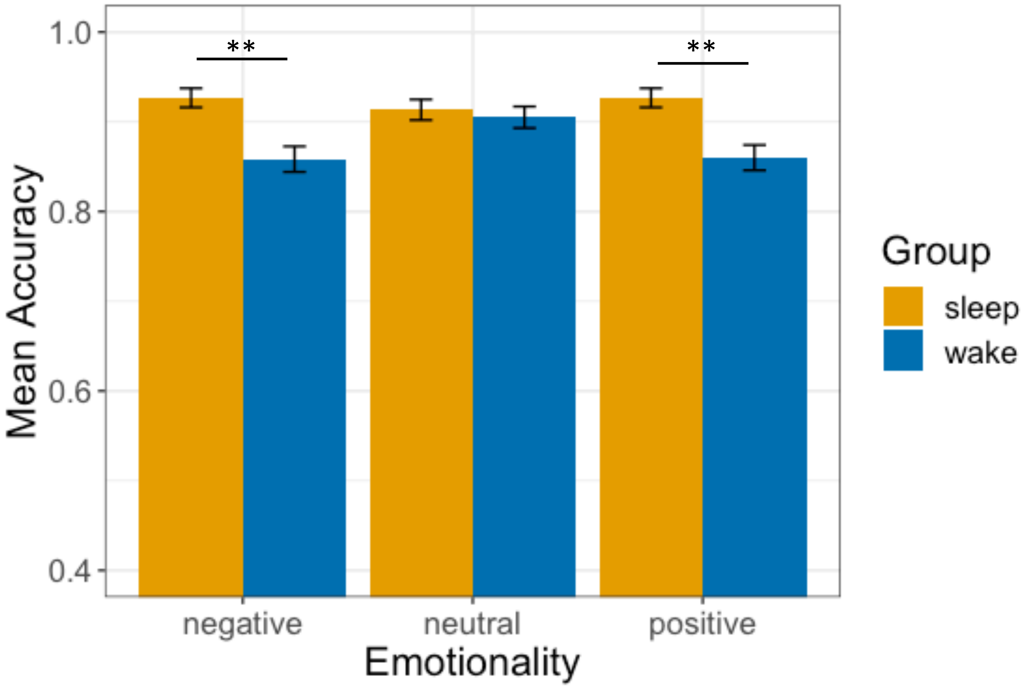


Figure 6.3. Performance accuracy in the sleep and wake groups for negative, neutral, and positive word pairs. Error bars show ±1 SEM, ***p* < .01.

3.3.2 Confidence Ratings

We asked participants to rate how confident they were in their responses to word pairs, with confidence ratings ranging from 1 (confident that the word pair was previously seen) to 6 (confident that the word pair was new). To analyse differences in confidence ratings, we first reversed participants ratings of unseen word pairs, so that a score of 1 would show high confidence in accurately accepting seen word pairs, but also accurately rejected unseen word pairs, and a score of 3 would show low confidence. This allowed us to more easily compare confidence ratings of seen and unseen word pairs.

Previously Seen Word Pairs

We ran Linear Mixed-Effects Models, first analysing confidence ratings to previously seen word pairs, to compare conventional to novel word pairs (contrast A), and then congruent versus incongruent word pairs (contrast B). See Figure 4 for means across all conditions.

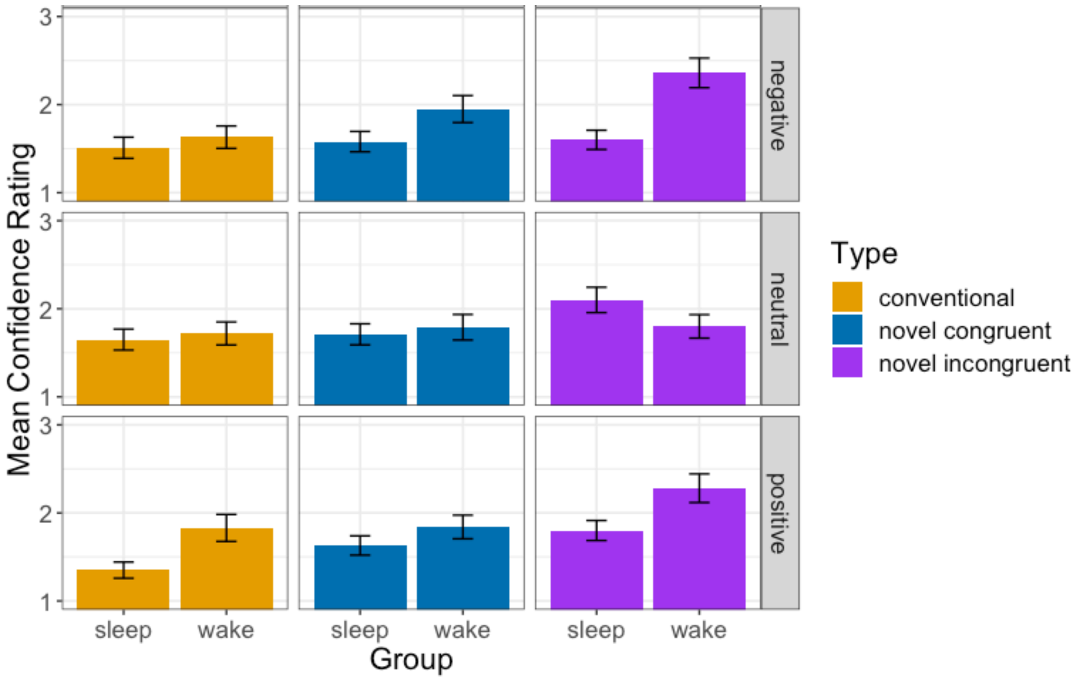


Figure 6.4. Mean confidence rating to previously seen word pairs for conventional, novel congruent, and novel incongruent metaphors, split by group and emotionality. Error bars show ±1 SEM.

Confirmatory Analyses

Adding the main effect of conventional versus novel word pairs to the model significantly improved model fit ($x^2(1) = 4.110, p < .043$). Conventional metaphorical word pairs were responded to with increased confidence novel word pairs (estimate = -.170, SE = .082, $t = -2.075$). Adding the main effect of congruent versus incongruent novel word pairs to the model also significantly improved model fit ($x^2(1) = 11.518, p < .001$). Congruent word pairs were responded to with increased confidence than incongruent word pairs (estimate = -.240, SE = .071, $t = -3.400$). Adding the main effect of group to the model led to a marginally significant effect, ($x^2(2) = 3.464, p = .063$). The wake group responded with less confidence than did the sleep group (estimate = .252, SE = .133, $t = 1.902$). Emotionality did not improve model fit ($x^2(2) = .029, p = .986$), nor did valence score at training ($x^2(1) = .172, p = .678$), or testing ($x^2(1) = .006, p = .937$), or arousal score at training ($x^2(1) = 3.022, p = .082$), or testing ($x^2(1) = 1.073, p = .300$).

Similarly to response accuracy, we also predicted that sleep would increase confidence for more difficult to learn word pairs, compared to the wake group. Thus we investigated the interactions between type of word pair and group. For contrast A (conventional versus novel word pairs), adding the interaction between group and word pair to the model did not improve model fit ($x^2(2) = 3.590, p = .166$). Adding the interaction between contrast B (congruent versus incongruent novel word pairs) and group also did not improve model fit ($x^2(2) = 3.901, p = .142$).

We also predicted that there would be an interaction between group and emotionality, such that the sleep group would show increased confidence for emotional word pairs than the wake group. Adding the interaction between group and emotionality to the model significantly improved model fit ($x^2(6) = 17.676$, $p = .007$). To further investigate this, we analysed the effect of emotionality on confidence ratings for the sleep and wake groups separately. For the sleep group, there was no significant difference in ratings of confidence between neutral and negative word pairs (estimate = .257, SE = .162, $t = 1.587$), neutral and positive word pairs (estimate = .227, SE = .162, $t = 1.402$), and positive and negative word pairs (estimate = .030, SE = .162, $t = .186$). For the wake group, there was again no difference in ratings of confidence between neutral and negative word pairs (estimate = -.210, SE = .170, $t = -1.234$), neutral and positive word pairs (estimate = -.213, SE = .170, $t = -1.254$), and positive and negative word pairs (estimate = .003, SE = .170, $t = .020$). We then analysed effects of sleep versus wake group on negative, positive and neutral word pairs separately. For negative word pairs, there was a significant effect of group, with the wake group reporting less confidence in their responses than the sleep group (estimate = .417, SE = .170, $t = 2.458$). For positive word pairs, the wake group again reported less confidence than the sleep group (estimate = .390, SE = .154, $t = 2.528$). For neutral word pairs, there was no difference in confidence ratings between the sleep and wake group (estimate = .050, SE = .137, $t = .364$).

We further predicted that ratings of valence and arousal would influence confidence ratings in the sleep group to a greater extent than the wake group. However, adding the interaction term group by valence at training ($x^2(4) = 4.740$, $p = .315$), arousal at training ($x^2(4) = 4.191$, $p = .242$), valence at testing ($x^2(4) = 4.391$, $p = .356$), and arousal at testing ($x^2(4) = 4.604$, $p = .330$) did not improve model fit.

Exploratory Analyses

We also analysed further interaction terms that were not part of our hypothesis. We added all possible 2-way and 3-way interactions one at a time to the model. Adding the interaction term novel versus conventional word pairs by arousal at training significantly improved model fit ($x^2(2) = 8.032$, $p < .018$). For novel word pairs, there was a significant effect of arousal at training (estimate = -.066, SE = .025, $t = -2.592$), confidence increased as arousal rating increased. There was no effect of arousal for conventional word pairs (estimate = .032, SE = .034, $t = .935$). This supports the findings from the accuracy data, and again suggests that the effect of emotionality on novel word pairs may have an immediate influence on learning and memory for novel information.

Seen versus Unseen Word Pairs

We then analysed whether there were any differences in confidence ratings between seen and unseen word pairs. Again we used the same random effects structure, and fixed effects of type of metaphor (novel, conventional), group (sleep, wake), emotionality (negative, positive, neutral), and seen versus unseen word pairs, plus all interaction terms. See Figure 5 for mean confidence rating across conditions.

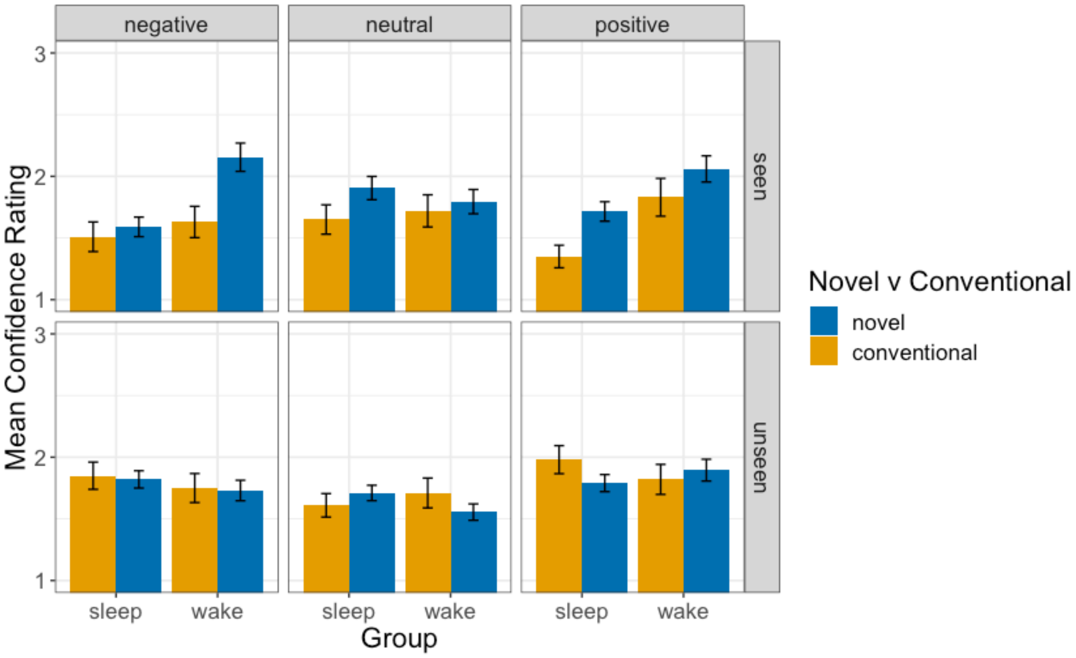


Figure 6.5. Mean confidence rating to seen and unseen word pairs for conventional versus novel metaphors, split by group and emotionality. Error bars show ±1 SEM.

Confirmatory Analyses

Adding the contrast between novel and conventional word pairs did not improve model fit ($x^2(1) = 1.981, p = .159$). The effect of group did not improve model fit ($x^2(1) = .811, p = .368$), nor did the main effect of seen versus unseen word pairs ($x^2(1) = .096, p = .757$) or emotionality ($x^2(2) = 1.420, p = .492$).

We then added interaction terms using a hypothesis-driven approach to our model building. We predicted that there would be an interaction between group and novel versus conventional word pairs, such that sleep would increase confidence for more difficult to learn stimuli. Adding this interaction to the model did not improve model fit ($x^2(3) = 2.890, p = .409$). We also predicted that there would be an interaction between group and emotionality. Adding the interaction term to the model improved model fit ($x^2(4) = 10.925, p = .027$). We therefore first investigated the effect of emotionality for the sleep and wake group separately. For the wake group, positive words pairs were responded to with less confidence than neutral word pairs (estimate = .238, SE = .104, t = 2.301). However, there was no significant difference in performance between negative and neutral word pairs (estimate = .170, SE = .104, t = 1.641) or negative and positive word pairs (estimate = -.068, SE = .104, t = 0.660). There was no effect of emotionality in the sleep group ($x^2(2) = .267$, $p = .874$). We then investigated the difference in confidence ratings between the sleep and wake group for negative, neutral, and positive word pairs separately. There was no significant difference between the sleep and wake group for negative (estimate = -.134, SE = .131, t = -1.025), positive (estimate = -.194, SE = .127, t = -1.524), or neutral word pairs (estimate = .076, SE = .112, t = 0.676).

Exploratory Analyses

We then added all other interaction terms one at a time into the model. Adding the interaction term group by seen versus unseen word pairs improved model fit ($x^2(3) = 16.975$, $p < .001$). To investigate the interaction further, we first investigated performance accuracy for seen versus unseen word pairs in the sleep and wake groups separately. For the wake group, seen word pairs were responded to with less confidence than unseen word pairs (estimate = .173, SE = .085, t = 2.033). There was no significant difference in performance between seen and unseen word pairs in the sleep group (estimate = .128, SE = .080, t = 1.593). We then investigated performance accuracy in the sleep versus wake groups for seen and unseen word pairs separately. There was no significant difference between the wake and sleep group for either previously seen word pairs (estimate = .196, SE = .135, t = 1.227), or unseen word pairs (estimate = -.049, SE = .151, t = -0.324).

The results further revealed a marginally significant 3-way interaction between group, emotionality, and seen versus unseen word pairs ($x^2(4) = 9.340$, $p = .053$; see Figure 6). We therefore investigated the emotionality by group interaction for seen and unseen word pairs separately. For seen word pairs, there was a significant group by emotionality interaction ($x^2(5) = 17.230$, $p = .004$). Analysis of the wake and sleep group separately revealed that adding emotionality to a random effects model did not significantly improve model fit for either the wake group ($x^2(2) = 2.017$, $p = .365$) or the sleep group ($x^2(2) = 2.915$, $p = .234$). However, for both negative and positive word pairs, the wake group was significantly less confident in their judgements than the sleep group (estimate = .417, SE = .170, $t = 2.458$; estimate = .390, SE = .154, $t = 2.528$). There was no difference in confidence between the sleep and wake group for neutral word pairs (estimate = -.050, SE = .137, $t = -0.364$). There was no significant group by emotionality interaction for unseen word pairs ($x^2(5) = 6.229$, $p = .285$).

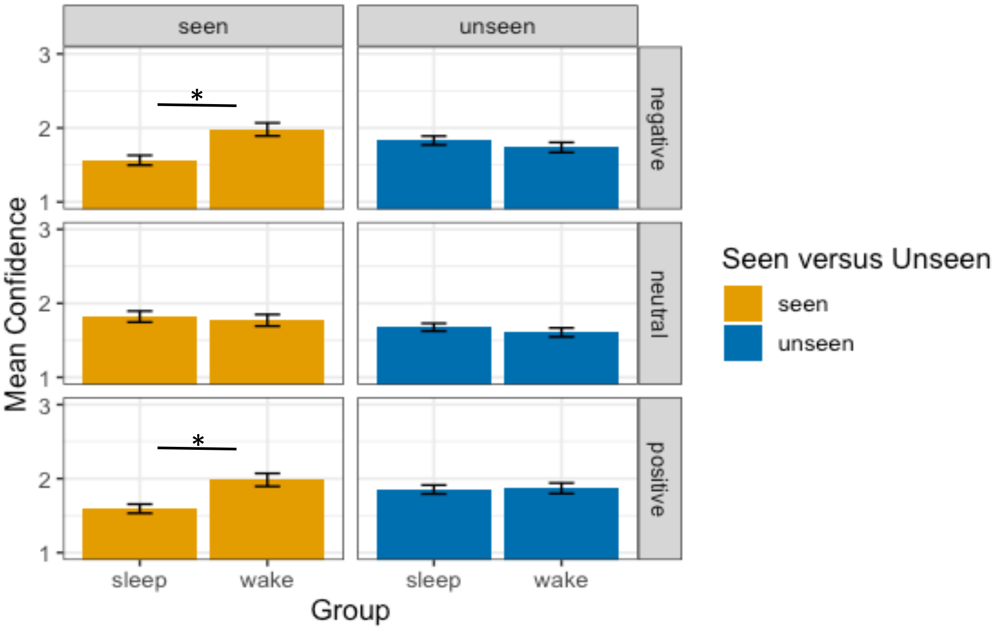


Figure 6.6. Mean confidence ratings for seen and unseen word pairs of negative, positive, and neutral valence, in the sleep and wake group. Error bars show ±1 SEM, * $t > 2$. A score of 1 indicates high confidence, whereas a score of 3 indicates low confidence.

3.3.3 Response Times

Response times (RT) for correct responses were analysed to test whether there were any speed-accuracy trade-off effects that may have influenced the accuracy findings, as well as to establish whether any factors significantly influenced participants' speed of access to stored information. Response times exceeding 2.5 standard deviations from the mean were removed from the analysis, leading to the removal of 2.439% of the data, and leaving 3157 data points. We reduced skew in the RT distribution by transforming latencies to $\log_{10}$(RT), and analysed the logRTs using Linear Mixed-Effects Modelling, considering the random effects of both participants and items on intercepts. We ran a series of models with the same random effects structure and varying in fixed effects structure. We compared models using likelihood ratio tests, using the REML = FALSE setting.

Previously Seen Word Pairs

Confirmatory Analyses

We analysed RT for correct recognition of those word pairs that participants were exposed to at training. See Figure 7 for means across conditions. Adding the main effect of novel versus conventional word pairs to the model did not improve model fit ($x^2(1) = 2.066, p = .151$). Similarly, the contrast between congruent versus incongruent novel word pairs did not improve model fit ($x^2(1) = 3.183, p = .074$). There was no effect of sleep versus wake group ($x^2(1) = .635, p = .426$), emotionality ($x^2(1) = .668, p = .716$), valence at training ($x^2(1) = 2.284, p = .131$) or testing ($x^2(1) = .138, p = .711$), or arousal at training ($x^2(1) = 2.153, p = .142$) or testing ($x^2(1) = .059, p = .808$).

Again, we also predicted that sleep would support learning of more difficult to learn word pairs compared to the wake group, and thus lead to reduced response times. We therefore investigated the interactions between type of word pair and group. For contrast A (conventional versus novel word pairs), adding the interaction between group and word pair to the model did not improve model fit ($x^2(3) = 2.909$, $p = .406$). Adding the interaction between contrast B (congruent versus incongruent novel word pairs) and group also did not improve model fit ($x^2(3) = 6.075$, $p = .108$).

We also predicted that there would be an interaction between group and emotionality, such that the sleep group would show quicker response times to emotional word pairs than the wake group. Adding the interaction between group and emotionality to the model did not improve model fit ($x^2(5) = 2.533$, $p = .772$). We further predicted that ratings of valence and arousal would influence response times in the sleep group to a greater extent than the wake group. However, adding the interaction term group by valence at training ($x^2(3) = 2.982$, $p = .395$), arousal at training ($x^2(3) = 2.822$, $p = .420$), valence at testing ($x^2(3) = 1.909$, $p = .592$), and arousal at testing ($x^2(3) = 1.084$, $p = .781$) did not improve model fit.

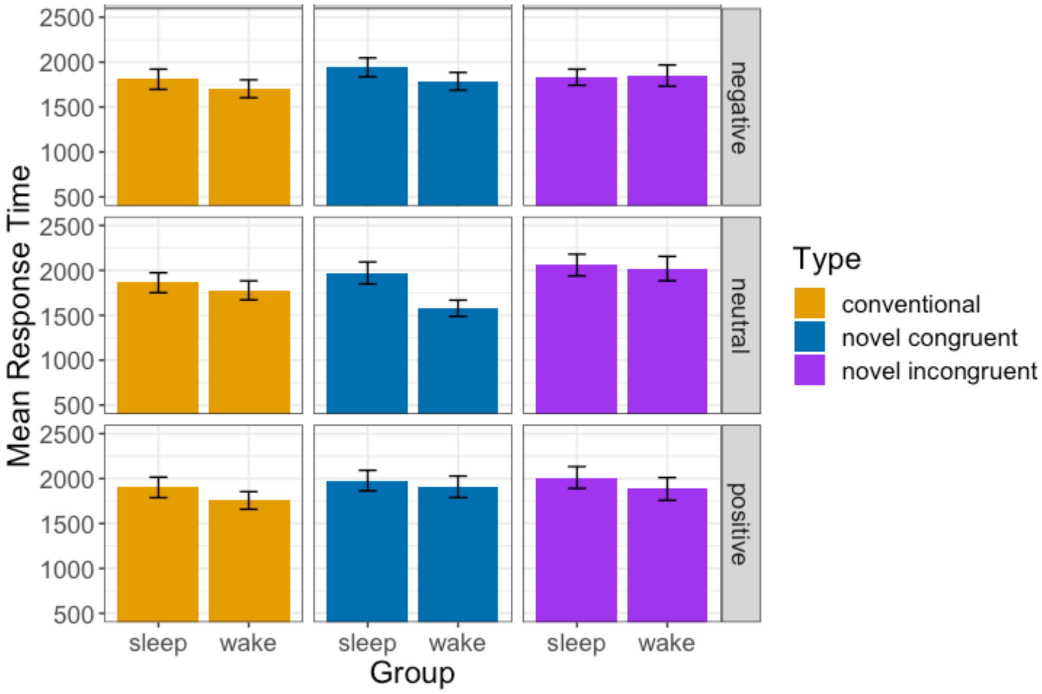


Figure 6.7. Mean response times for conventional, novel congruent, and novel incongruent seen word pairs in the sleep and wake group, split by emotionality. Error bars show ±1 SEM.

Seen versus Unseen Word Pairs

Confirmatory Analyses

We then analysed RT for correct responses to both seen and unseen word pairs. See Figure 8 for means across conditions. We first added all main effects to the model. The main effect of novel versus conventional word pairs did not improve model fit compared to a model with only random effects ($x^2(1) = .823$, $p = .364$). There was also no significant effect of group ($x^2(1) = 2.307$, $p = .129$), or emotionality ($x^2(2) = 1.576$, $p = .455$). Adding the main effect of seen versus unseen word pairs to a model with only random effects significantly improved model fit, ($x^2(1) = 14.008$, $p < .001$). As predicted, previously unseen word pairs were responded to significantly slower than previously seen word pairs (estimate = .043, SE = .011, $t = 3.901$).

We then added interaction terms again using a hypothesis-driven approach to model building. We predicted that there would be an interaction between group and novel versus conventional word pairs, such that sleep would reduce response times for more difficult to learn stimuli. Adding this interaction to the model did not improve model fit ($x^2(3) = 4.465$, $p = .215$). We also predicted that there would be an interaction between group and emotionality. Adding the interaction term to the model did not improved model fit ($x^2(3) = 3.354$, $p = .500$). = .874).

Exploratory Analyses

We then added all other interaction effects to the model as exploratory analysis. The analysis revealed a significant interaction between seen versus unseen word pairs and group ($x^2(2) = 12.995$, $p = .002$). First we investigated any difference between the sleep and wake group for seen and unseen word pairs separately. For previously seen word pairs, there was no significant difference between groups (estimate = -.023, SE = .030, $t = -0.800$). For previously unseen word pairs, there was a marginally significant difference between groups; the wake group responded quicker than the sleep group (estimate = -.064, SE = .034, $t = -1.891$). We then investigated the difference in response time to previously seen and unseen word pairs in the sleep and wake group separately. In the sleep group, participants took significantly longer to respond to unseen than seen word pairs (estimate = .065, SE = .013, $t = 4.827$). For the wake group there was a marginally significant difference; unseen word pairs were responded to slower than seen word pairs (estimate = .022, SE = .013, $t = 1.704$).

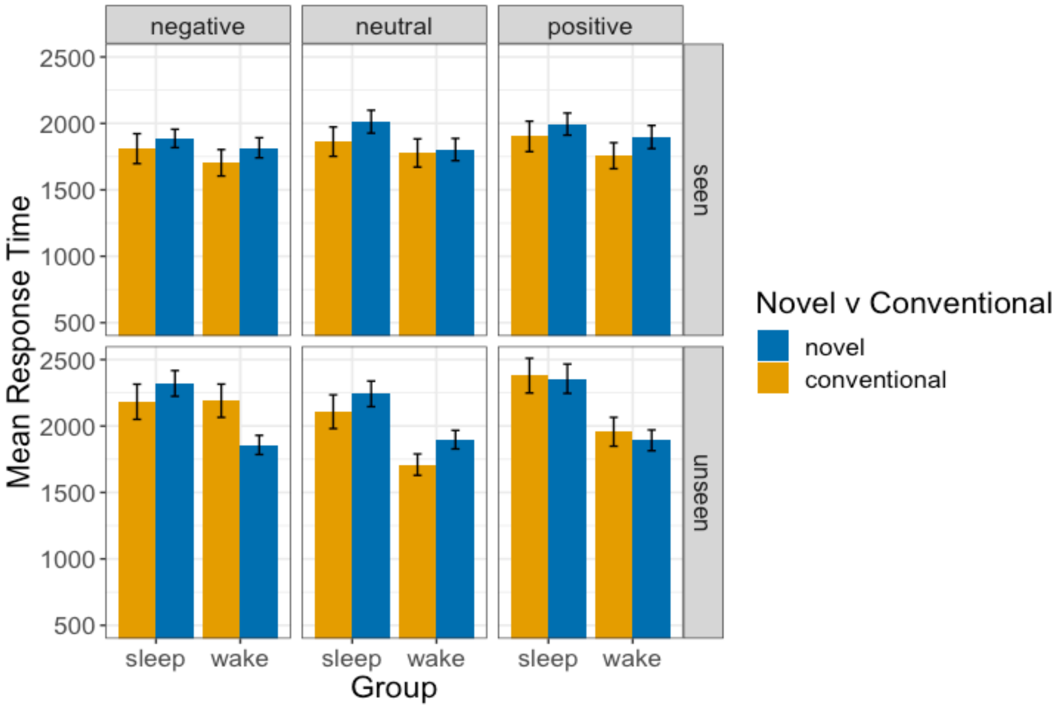


Figure 6.8. Mean response times for seen and unseen novel and conventional word pairs in the sleep and wake group, split by emotionality. Error bars show ±1 SEM.

3.4 Time of Day Control

During the training session we measured participants' ability to recall definitions for a second time, to gain a measure of recall accuracy after training. This allowed us to investigate whether time of day of training had any effect on recall performance accuracy. Since half of participants were exposed to the word pairs and definitions in the morning, and the other half in the evening, we may see differences in performance due to time of day. Previous research suggests increases in memory performance in the morning than in the evening (Schmidt, Collette, Cajochen, & Peigneux, 2007), thus we may expect to see increases in recall performance accuracy immediately after training in the wake than the sleep group.

Although all participants provided definitions for a second time during training, due to computer error only half of the data was collected (10 participants in the wake group, 10 in the sleep group). We thus ran General Linear Mixed-Effects Models on data from these 20 participants, again adding the random effects of participants and items, and adding each main effect and interaction term to the model. Critically, we found no effect of sleep versus wake group, ($x^2(1) = .586$ $p = .444$). Thus, accurate recall of definitions did not differ as a result of learning in the morning or the evening.

Indeed, the results replicated those found at testing; adding the main effects of type of metaphor and arousal at testing significantly improved model fit. For type of metaphor, ($x^2(2) = 78.977$ $p = 2.2 \text{ x } 10^{16}$), the definitions of novel incongruent metaphorical word pairs were recalled significantly less accurately than both conventional word pairs (estimate = -4.795, SE = .431, z = -11.113, $p = 2 \text{ x } 10^{16}$), and novel congruent word pairs (estimate = 4.400, SE = .417, z = -10.542, $p = 2 \text{ x } 10^{16}$). Accurate recall of definitions of conventional and novel congruent word pairs did not significantly differ (estimate = 0.395, SE = .354, z = 1.114, $p = .265$). In the same way as the analysis of recall accuracy at testing, we then added main effects of emotionality (negative, positive, and neutral) and ratings of valence and arousal to the model. We again found that participants' ratings of arousal at testing significantly improved model fit, ($x^2(1) = 8.316$, $p = .003$), with increased ratings of arousal leading to greater performance accuracy (estimate = .230, SE = .081, z = 2.852, $p = .004$). No other main effects or interactions were significant.

3.5 Ratings of Emotionality

We analysed differences in participants ratings of valence and arousal at training compared to at testing (see Table 11 for overall means and SD's as well as split by sleep and wake group). A paired samples t-test indicated no significant difference in valence scores at training versus testing (*t* = -.888, *p* = .375), however there was a significant difference between arousal scores, (*t* = 2.088, *p* = .037). Arousal scores were higher at training than testing. However, despite significance, mean arousal scores at training and testing only differed by a rating of .081 (see Table 9).

If sleep reduces the affective ratings of word pairs (van der Helm & Walker, 2009), we would expect the sleep group to report reduced ratings of valence and arousal at testing compared to at training. Thus, we ran further t-tests for the sleep and wake group separately. For the sleep group, there was no significant difference in valence scores at training versus testing, (*t* = -.740, *p* = .459). However, ratings of arousal were higher at training than at testing (*t* = 2.834, *p* = .005). For the wake group, there was no significant difference at training versus testing for either ratings of valence (*t* = -.559, *p* = .576), or of arousal (*t* = .337, *p* = .736).

Table 6.9

Means and standard deviations for ratings of valence and arousal at training and testing.

	Valence		Arousal	
	Training	Testing	Training	Testing
Overall	3.974(2.031)	4.006(1.892)	4.387(1.639)	4.306(1.456)
Sleep group	3.980(2.048)	4.015(1.713)	4.328(1.626)	4.170(1.426)
Wake group	3.970(2.019)	3.999(1.976)	4.435(1.650)	4.417(1.471)

4. Discussion

The present study examined the role of sleep on the learning of and memory for both novel and conventional metaphorical word pairs varying in emotionality. First, we aimed to replicate the large body of research indicating an overall benefit of sleep to both learning and memory (e.g., Born, Rasch, & Gais, 2006; Dumay & Gaskell, 2007; Rasch & Born, 2013). We found no significant main effect of sleep versus wake group for recall of the definitions, however as predicted, the sleep group responded significantly more accurately than the wake group in the recognition task. We further predicted that more difficult to remember definitions would result in a greater difference in performance between the sleep and wake group (Sio et al., 2013; Stickgold & Walker, 2004), however this was not supported by the present study; we found no difference between sleep and wake groups in recall of definitions for either incongruent or congruent metaphorical word pairs. We suggest that this lack of an effect of sleep compared to wake on memory for definitions may be due to a floor effect for the novel incongruent metaphorical word pairs. Participants on average remembered only .06% of incongruent metaphors, suggesting that the task was too difficult.

Another main aim of the present study was to investigate what effect emotionality has on the learning of and memory for metaphorical word pairs. We gained measures of both valence (from negative to positive) and arousal (from calm to exciting/agitating), to determine whether specific affective ratings influenced performance accuracy. We also investigated whether categories of both negative and positive word pairs were remembered better than those word pairs categorised as non-emotional. Based on previous research, we hypothesised that emotionality would lead to both increased learning of novel metaphorical word pairs and increased accurate recognition of to-be-remembered word pairs (Adelman & Estes, 2013; Kensinger & Corkin, 2004; McGaugh, 2004). However, due to inconsistencies within previous literature, we could not predict what specific effects valence or arousal may have. Kensinger and Corkin (2003) found that increased valence improved memory performance, but this was strongest when those words were higher in arousal, whereas Adelmann and Estes (2013) found an effect of valence independent of arousal.

For recall of definitions, we found that ratings of arousal at testing significantly correlated with performance accuracy. The higher in arousal that word pairs were rated, the more likely that participants were able to accurately recall the definitions. Similarly, for accurate recognition of to-be-remembered word pairs, we found a significant effect of arousal rating at training for novel word pairs, with increased ratings of arousal leading to greater response accuracy. This effect of arousal was further evident in ratings of confidence, with increased ratings of arousal leading to increased confidence of recognition accuracy. These findings help to clarify effects of emotionality on memory, suggesting that increases in memory for emotional information may be specific to arousal, supporting previous suggestions of arousal, rather than valence, influencing memory performance (Kensinger & Corkin, 2003; Mather & Sutherland, 2009).

We also aimed to investigate the role of sleep in both emotional and non-emotional learning and memory. Based on previous findings, we hypothesised that sleep would boost memory for emotional compared to non-emotional word pairs (van der Helm & Walker, 2009; Walker, 2009). The results of the current study support these findings; performance accuracy in the recognition task was greater in the sleep group than the wake group for both negative and positive word pairs, whereas no difference between the sleep and wake group was found for neutral word pairs. These findings were further supported by increased recognition confidence for negative and positive word pairs in the sleep than wake group specifically for those word pairs that were previously seen at training.

The present study therefore indicates that word pairs of both negative and positive valence were remembered more accurately after sleep than wake. The results support previous findings of sleep aiding consolidation of negative emotional information (e.g., Hu et al., 2006; Nishida et al., 2009; Payne et al., 2008; Wagner et al., 2005), and extends these findings to include positive emotional memories. The present study is thus one of the first to investigate the effects of both negative *and* positive emotion in comparison to non-emotional information in sleep-related memory consolidation. Together with our findings of increased ratings of arousal supporting performance accuracy, these results suggest an overall benefit of arousing information of both negative and positive valence on memory performance, and this benefit is further enhanced by the offline consolidation processes that occur during sleep.

Based on the sleep to forget, sleep to remember hypothesis (van der Helm & Walker, 2009), we predicted that not only would the current study provide support for sleeping to remember emotional information, but also that those in the sleep group would show a reduction in their affective ratings of the word pairs. The results of the current study partially support a “sleep to forget” hypothesis. The sleep group rated word pairs as significantly less arousing after a night of sleep, whereas no difference in affective ratings was evident in the wake group. However, the mean reduction in ratings of arousal after sleep was extremely small and did not extend to ratings of valence. One possible explanation for only a very small effect of sleep on reducing the affective tone of the word pairs is that the SFSR hypothesis suggests the affective tone of a memory is reduced over time, and over a number of nights of sleep. Indeed, some previous findings do not support a “sleep to forget” hypothesis after only one night of sleep; Wiesner et al., (2015) found that overnight sleep did not reduce ratings of arousal of pictures, and Wagner, Fischer and Born (2002) found that one night of sleep led to an increase in ratings of arousal. Thus, further research is necessary to establish whether the affective tone of emotional memories can be reduced after a number of nights of sleep related memory consolidation.

The findings of the current study therefore support previous suggestions of increasing arousal leading to greater memory performance after sleep for images (Hu et al., 2006; Nishida et al., 2009), texts (Wagner et al., 2001), and objects in scenes (Payne et al., 2008) and extends these findings to conventional and novel metaphorical word pairs. Previous studies of emotion and sleep have focused on stimuli that already contain an emotional response for the participants (e.g., Hu et al., 2006, Wagner et al., 2001), whereas the present study sheds light on the effects of emotion on memory when the meaning of the word is not known but is learned by convention. This study is one of the first to suggest that emotionality can influence learning and memory even when meaning is not captured in the information.

Furthermore, the present study indicates that the effects of emotion and sleep are similar regardless of whether the emotional stimuli are previously known or novel. We found that effects of emotion on metaphorical word pairs is immediate, irrespective of whether emotionality is already known (as in the conventional metaphorical word pairs) or is learned (as in the novel word pairs). Since metaphorical expressions have previously been found to show increased emotional brain responses compared to literal expressions (Citron & Goldberg, 2014), this may explain why influences of emotionality are evident even for novel word pairs. Furthermore, we found that sleep enhanced emotional memory regardless of whether the word pairs were previously known or novel, again supporting the suggestion that effects of emotion on memory occur immediately after learning.

However, the results of the present study require further investigation with an increased sample size, to further enhance power of the effects found. Further, the multiple comparisons conducted to investigate the effects of sleep compared to wake on emotional memory lead to the potential for increased risk of type one errors. Although our findings support previous evidence of an effect of sleep on emotional memory performance, further research should aim to replicate these findings with increased power, to further understand the effects of sleep on the learning of and memory for emotional metaphorical word pairs.

This study is one of the first to investigate effects of both negative, positive, and neutral emotion in sleep-related memory consolidation, and is the first to suggest that emotionality can influence learning and memory regardless of whether the meaning of the word pairs is known. The findings suggest a boost in memory for figurative language higher in arousal, and indicates that sleep specifically supports memory for figurative language of both negative and positive valence.

References

Adelman, J. S., & Estes, Z. (2013). Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition, 129(3),* 530-535. doi:10.1016/j.cognition.2013.08.014

Bambini, V., Gentili, C., Ricciardi, E., Bertinetto, P. M., & Pietrini, P. (2011). Decomposing metaphor processing at the cognitive and neural level through functional magnetic resonance imaging. *Brain Research Bulletin, 86*, 203-216. doi:10.1016/j.brainresbull.2011.07.015

Born, J., Rasch, B., & Gais, S. (2006). Sleep to Remember. *The Neuroscientist, 12*(5), 410-424. doi:10.117/1073858406292647

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49-59. doi:10.1016/0005-7916(94)90063-9

Cahill, L., Haier, R., Fallon, J., Alkire, M., Tang, C., Keator, D., . . . Mcgaugh, J. (1996). Amygdala activity at encoding correlated with long-term, free recall of emotional information. *Proceedings of the National Academy of Sciences of the United States of America, 93*(15), 8016-8021. doi:10.1073/pnas.93.15.8016

Cardillo, E. R., Watson, C. E., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2012). From novel to familiar: Tuning the brain for metaphors. *NeuroImage, 59*(4), 3212-3221. doi:10.1016/j.neuroimage.2011.11.079

Carter, R. (2004). *Language and creativity [electronic resource] : The art of common talk*. London ; New York: Routledge.

Citron, F. M. M., & Goldberg, A. E. (2014). Metaphorical sentences are more emotionally engaging than their literal counterparts. *Journal of Cognitive Neuroscience, 26*(11), 2585-2595. doi:10.1162/jocn_a_00654

Citron, F. M. M., Güsten, F., Michaelis, N., & Goldberg, A. E. (2016). Conventional metaphors in longer passages evoke affective brain response. *NeuroImage, 139*, 218-230. doi:10.1016/j.neuroimage.2016.06.020

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience, 11*(2), 114-126. doi:10.1038/nrn2762

Diekelmann, S., Landolt, H., Lahl, O., Born, J., & Wagner, U. (2008). Sleep loss produces false memories. *PloS one, 3*(10), e3512-e3512. doi:10.1371/journal.pone.0003512

Dumay, N., & Gaskell, G. (2007). Sleep-Associated Changes in the Mental Representation of Spoken Words. *Psychological Science, 18*(1), 35-39. doi:10.1111/j.1467-9280.2007.01845.x

Fainsilber, L., & Ortony. A. (1987). Metaphorical uses of language in the expression of emotions. *Metaphor and Symbolic Activity, 2*(4), 239-250. doi:10.1207/s15327868ms0204_2

Forgács, B., Lukács, Á., & Pléh, C. (2014). Lateralized processing of novel metaphors: Disentangling figurativeness and novelty. *Neuropsychologia, 56*, 101-109.

Giora, R. (1999). On the priority of salient meanings: Studies of literal and figurative language. *Journal of Pragmatics, 31*(7), 919-929. doi:10.1016/S0378-2166(98)00100-3

Glucksberg, S. (1998). Understanding Metaphors. *Current Directions in Psychological Science, 7*(2), 39-43. 10.1111/1467-8721.ep13175582

Groch, S., Wilhelm, I., Diekelmann, S., & Born, J. (2013). The role of REM sleep in the processing of emotional memories: Evidence from behavior and event-related potentials. *Neurobiology of Learning and Memory, 99*, 1-9. doi:10.1016/j.nlm.2012.10.006

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. (1973). Quantification of Sleepiness: A New Approach. *Psychophysiology, 10*(4), 431-436.

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

Ito, T., Cacioppo, J., & Lang, P. (1998). Eliciting Affect Using the International Affective Picture System: Trajectories through Evaluative Space. *Personality and Social Psychology Bulletin, 24*(8), 855-879. doi:10.1177/0146167298248006

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition, 31*(8), 1169-1180. doi:10.3758/BF03195800

Kensinger, E. A., & Corkin, S. (2004). Two routes to emotional memory: distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences of the United States of America, 101*(9), 3310-3315. doi:10.1073/pnas.0306408101

Lakoff, G., & Johnson, Mark. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Labar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience, 7*(1), 54-64. doi:10.1038/nrn1825

Liu, P. Q., Connell, L., & Lynott, D. (2018). Ease-of-Simulation Norms for 452 Adjective Metaphors. *Manuscript under Review*, 1–27.

Marshall, L., & Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends in Cognitive Sciences, 11*(10), 442-450. doi:10.1016/j.tics.2007.09.001

Mashal, N., & Itkes, O. (2013). The effects of emotional valence on hemispheric processing of metaphoric word pairs. *Laterality: Asymmetries of Body, Brain and Cognition*, *19*(5), 511-521. doi: 10.1080/1357650X.2013.862539

Mather, M., & Sutherland, M. (2009). Disentangling the effects of arousal and valence on memory for intrinsic details. *Emotion Review, 1*(2), 118-119. doi:10.1177/1754073908100435

McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience, 27*(1), 1-28. doi:10.1146/annurev.neuro.27.070203.144157

Nishida, M., Pearsall, J., Buckner, R. L., & Walker, M. P. (2009). REM sleep, prefrontal theta, and the consolidation of human emotional memory. *Cerebral Cortex, 19*(5), 1158-1166. doi:10.1093/cercor/bhn155

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

Plihal, W., & Born, J. (1997). Effects of Early and Late Nocturnal Sleep on Declarative and Procedural Memory. *Journal of Cognitive Neuroscience, 9*(4), 534-547. doi:10.1162/jocn.1997.9.4.534

Pynte, J., Besson, M., Robichon, F., & Poli, J. (1996). The Time-Course of Metaphor Comprehension: An Event-Related Potential Study. *Brain and Language, 55*(3), 293-316. doi:10.1006/brln.1996.0107

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

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

Reisenzein, R. (1994). Pleasure-Arousal Theory and the Intensity of Emotions. *Journal of Personality and Social Psychology,67*(3), 525-539. doi:10.1037/0022-3514.67.3.525

Schmidt, G. L., & Seger, C. A. (2009). Neural correlates of metaphor processing: The roles of figurativeness, familiarity and difficulty. *Brain and Cognition, 71*(3), 375-386.

Sio, U., Monaghan, P., & Ormerod, T. (2013). Sleep on it, but only if it is difficult: Effects of sleep on problem solving. *Memory & Cognition, 41*(2), 159-166. doi:10.3758/s13421-012-0256-7

Stickgold, R., & Walker, M. (2004). To sleep, perchance to gain creative insight? *Trends in Cognitive Sciences, 8*(5), 191-192. doi:10.1016/j.tics.2004.03.003

van der Helm, E., Yao, J., Dutt, S., Rao, V., Saletin, J., & Walker, M. (2011). REM sleep depotentiates amygdala activity to previous emotional experiences. *Current Biology, 21*(23), 2029-2032. doi:10.1016/j.cub.2011.10.052

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

Wagner, U., Fischer, S., & Born, J. (2002). Changes in emotional responses to aversive pictures across periods rich in slow-wave sleep versus rapid eye movement sleep. *Psychosomatic Medicine, 64*(4), 627-634.

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*(2), 112-9. doi:10.1101/lm.36801

Walker, M. (2009). The Role of Sleep in Cognition and Emotion. *Annals of the New York Academy of Sciences, 11561*(1), 168-197. Doi:10.111/j.1749-6632.2009.04416.x

Wiesner, C. D., Pulst, J., Krause, F., Elsner, M., Baving, L., Pedersen, A., . . . Göder, R. (2015). The effect of selective REM-sleep deprivation on the consolidation and affective evaluation of emotional memories. *Neurobiology of Learning and Memory, 122*, 131-141. doi:10.1016/j.nlm.2015.02.008

Wilson, J. K., Baran, B., Pace-Schott, E. F., Ivry, R. B., & Spencer, R. M. C. (2012). Sleep modulates word-pair learning but not motor sequence learning in healthy older adults. *Neurobiology of Aging, 33*(5), 991-1000. doi:10.1016/j.neurobiolaging.2011.06.029

7. General Discussion

In this thesis, I have explored the role of sleep in veridical and false memory consolidation, using a range of behavioural techniques as well as polysomnography (PSG) to understand the factors that influence the integration of memories into neocortical systems during offline periods of sleep. Previous literature on sleep and memory processing provide insights into the role of sleep on veridical declarative memory, suggesting that sleep supports the reactivation of memories encoded in hippocampal systems, leading to the integration of memories into neocortical networks (Marshall & Born, 2007; Rasch & Born, 2013). However, this interplay between hippocampal and neocortical systems leads to a qualitative change in the memory representation, such that the memory becomes more gist-based and independent of context. Furthermore, sleep has been found to selectively reactivate and consolidate certain memories, in particular those of increased emotionality (Rasch & Born, 2013; van der Helm & Walker, 2009). The aim of the current thesis was thus to outline the current literature on sleep and veridical and false memories, and to expand this to understand how various factors, namely lateralisation and emotionality, influence the consolidation of memories.

7.1 Summary of Results

The literature surrounding sleep and false memories indicates potential moderating variables that may influence whether broad semantic associations are activated in memory. In Chapter 2, we conducted a meta-analysis to first establish the overall effect size associated with sleep-dependent memories within the DRM paradigm, and to establish any task-specific influences on veridical and false memories. Overall, sleep did not increase veridical or false memory compared to wakefulness. However, the results help to clarify previous mixed results, indicating that recall but not recognition testing, and shorter list lengths, enhance sleep-dependent veridical and false memories.

During sleep the reactivation and integration of memories into neocortical systems supports increased veridical memories in the left hemisphere (Monaghan, Shaw, Ashworth-Lord, & Newbury, 2017), and may lead to increased false memories in the right hemisphere (Shaw & Monaghan, 2017). Chapter 3 aimed to explore whether effects of lateralisation at encoding, rather than at retrieval, have a similar influence on subsequent veridical and false memories. We used the same materials as in Monaghan et al. (2017), and an eye-tracker was used to control for participants' eye-movements away from the central fixation point, ensuring that DRM word lists were presented to only the left or right visual field. We found no effect of lateralisation at encoding on either veridical or false memories, at both immediate testing and after a period of sleep versus wakefulness. Thus, hemispheric processing of memories at encoding was not found to influence subsequent retrieval, and effects of lateralisation on sleep-dependent memories may instead be specific to retrieval processes.

Chapter 4 investigated the effect of sleep on emotional versus neutral DRM memories. The previous literature investigating selective enhancement of emotional memories does not disentangle differences between negative and positive emotion (e.g., McKeon, Pace-Schott & Spencer, 2012; Nishida, Pearsall, Buckner, & Walker, 2009; Payne, Stickgold, Swanberg, & Kensinger, 2008), thus we investigated the role of sleep on veridical and false memories for negative, neutral, *and* positive DRM word lists. As predicted, we found an overall effect of emotionality; both negative and positive information led to increased veridical and false memories. Interestingly, the results indicated that a night of sleep increased memory for negative compared to neutral veridical memories, whereas wake led to increases in positive veridical and false memories.

Chapter 5 extended the findings of Chapter 4 to investigate specific sleep architecture involved in the processing of emotional memories. Using the same DRM word lists as in Chapter 4, we compared performance between those who took a daytime nap and those who remained awake. We found that sleep improved veridical memory for emotional compared to neutral word lists, however there was no increase in memory for unseen, related information. Furthermore, using polysomnography we found that sleep spindles occurring during slow wave sleep (SWS) related to veridical memory performance for emotional information; increased sleep spindle density was correlated with increased veridical memory for emotional word lists.

Finally, Chapter 6 explored the role of sleep in learning novel figurative language and on affective ratings of to-be-remembered information. We investigated memory for negative, positive, and neutral metaphorical word pairs, and collected affective ratings of arousal and valence for each word pair. We found that higher ratings of arousal correlated with increased recall and recognition of metaphorical word pairs. Furthermore, we found that overnight sleep increased accurate recognition of emotional word pairs compared to the same time spent awake.

7.2 Integration of Results and Implications for the Literature

7.2.1 Recognition Testing

The studies employing the DRM paradigm within this thesis (Chapters 3, 4, & 5) used recognition testing as a measure of veridical and false memory. There were a number of methodological reasons for this. Firstly, In Chapter 3 we investigated the effects of lateralisation at encoding on veridical and false memories. Since previous findings of lateralisation effects at retrieval used recognition testing (e.g., Monaghan et al., 2017; Shaw & Monaghan, 2017), we followed a similar methodology to allow for direct comparisons between lateralisation effects at encoding versus at retrieval. Second, in Chapters 4 and 5 we investigated sleep-dependent veridical and false memories for emotional versus neutral DRM word lists. Previous results of an effect of emotionality on false memories without sleep suggest a greater influence of emotionality during tests of recognition as opposed to recall (El Sharkawy, Groth, Vetter, Beraldi, & Fast, 2008; Howe, Candel, Otgaar, Malone, & Wimmer, 2010), with increased false recognition of negative compared to neutral words (Brainerd, Stein, Rohenkohl, & Reyna, 2008; El Sharkaway et al., 2008; Howe et al., 2010), and a decrease in false recognition of positive than neutral words (Brainerd et al., 2008). However, there are also suggestions that emotionality may increase false memories in tests of recall (Bauer, Olheiser, Altarriba, & Landi, 2009). Since effects of emotion on recall performance are unclear, we therefore chose to employ tests of recognition to investigate emotional veridical and false memories, to enable us to make clearer predictions as to the effect of sleep on emotional memory. Furthermore, previous research of effects of sleep on emotional memory using the DRM paradigm has focused on recall testing (McKeon, Pace-Schott, & Spencer, 2012), with increases in veridical memory for neutral but not negative words after sleep, and no difference in false recall of words associated with negative versus neutral lists. These findings therefore support suggestions of a lack of an effect of emotion on false recall performance. Consequently, another reason that we chose recognition testing in the present studies was to compare effects of sleep on emotional veridical and false recognition to those found by McKeon et al. when employing a test of recall.

In Chapter 2 we presented a meta-analysis to investigate the overall effect of sleep versus wake on DRM memory performance, as well as to establish whether any moderating factors influence this effect. Previous studies of effects of sleep on DRM veridical and false memories without emotion provide inconsistent results dependent on whether participants are tested on their recall or recognition accuracy. Findings from studies employing free recall indicate an increase in both accurate and false memories (Diekelmann, Born, & Wagner, 2010; Payne et al., 2009), whereas tests of recognition have been found to enhance, have no effect, or reduce false memories (Diekelmann, Landolt, Lahl, Born, & Wagner, 2008; Fenn, Gallo, Margoliash, Roediger, & Nusbaum, 2009; Monaghan et al., 2017). We hypothesised that tests of recall would lead to enhanced false memories, whereas tests of recognition would provide protection against false memories. This was supported by the findings, revealing no effect of sleep on either veridical or false recognition.

The results of the meta-analysis thus support suggestions of a spreading activation theory of sleep and memory (Gallo & Roediger, 2003; Howe, Wimmer, Gagnon, & Plumpton, 2009; Roediger, Watson, McDermott, & Gallo, 2001). During tests of recognition, unseen, related words are supressed due to monitoring cues, whereas during free recall these monitoring cues are not as easily available, leading to greater activation of associated words (activation/monitoring framework; Collins & Loftus, 1975). Sleep is suggested to further enhance source monitoring (Johnson, Hashtroudi, & Lindsay, 1993), providing one explanation for the difference in sleep-dependent false memory effects between recall and recognition testing. The results of the empirical studies presented in this thesis further support the meta-analysis; studies of DRM memory performance (Chapters 3, 4, & 5) reveal no overall effect of sleep on either veridical or false recognition.

7.2.2 Hemispheric Processing of Memories

Chapter 3 aimed to investigate whether encoding word lists in the left or right hemisphere influenced subsequent retreival of the memories. Previous studies of lateralisation effects have found a LH advantage for language processing (Ellis, 2004; Pirozzolo & Rayner, 1977), and a RH advantage for priming of distantly related words (Beeman, Friedman, Grafman, Perez, Diamond, & Lindsay, 1994; Chiarello, Burgess, Richards, & Pollock, 1990). Thus, the fine versus coarse semantic coding hypothesis suggests that the LH is able to activate close semantic networks of related words, whereas the RH activates broad semantic networks of distantly related words (Beeman & Bowden, 2000; Beeman & Chiarello, 1998; Burgess & Simpson, 1988; Monaghan, Shillcock, & McDonald, 2004). Previous findings of DRM false memories support this, with results indicating increased veridical memory for words presented to the LH, and increased false memory for words presented to the RH (Bellamy & Shillcock, 2007; Faust, Ben-Artzi, & Harel, 2008; Ito, 2001; Westerberg & Marsolek, 2003). The first experiment of Chapter 3 examined whether encoding words in the RH or LH led to the same effect as previously found during retrieval. We found that encoding word lists in the LH or RH had no effect on subsequent retrieval processes; there was no difference in veridical or false memories dependent on lateralisation. This supports previous findings that suggest no hemispheric differences in hippocampal activation during initial encoding of stimuli (Hocking, McMahon & De Zubicaray, 2009).

Chapter 3 also aimed to investigate the role of sleep on the encoding of word lists in the LH or RH. Although we found no effect of lateralisation at encoding on immediate recognition performance, the way that memories are encoded in the hippocampus may influence how they are reactivated and integrated into the long-term store in the neocortex. Previous findings of sleep and hemispheric processing at retrieval have found a LH bias for veridical memories (Monaghan et al., 2017), and a RH bias for false memories (Shaw & Monaghan, 2017) after sleep compared to wake. These studies support a spreading activation account of sleep-dependent memory consolidation; sleep supports veridical memory consolidation in the LH coarse semantic networks, but also increases the number of semantically associated words that are activated in the RH broad semantic networks. In Experiment 2, encoding word lists in the LH or RH had no effect on sleep-dependent memory consolidation. Thus, whether hippocampal-based memories are encoded in the LH or RH has no effect on consolidation processes, and supports an equal spread of activation of memories within the LH and RH. We conclude that effects of lateralisation on sleep-dependent memory consolidation are likely to be specific to the processes of retrieval, when memories are either able to be accurately retreived from fine semantic networks of the LH, or related words are falsely activated in the broad semantic networks of the RH.

7.2.3 The Role of Sleep in Emotional Memory Consolidation

Theories of memory consolidation suggest that memories are initially represented in hippocampal systems, where they are encoded with their specific detail. During the processes of consolidation, memories are reactivated and integrated into neocortical systems that store gist-based memories with broader associations (McClelland, McNaughton & O'Reilly, 1995; Kumaran, Hassabis, & McClelland, 2016; Nadel & Moscovitch, 1997). However, not all memories are remembered equally, with evidence suggesting a bias for the consolidation of emotional memories (van der Helm & Walker, 2009; Walker, 2009). Sleep in particular is suggested to enhance this selective consolidation of emotional memories, with a specific influence of rapid-eye-movement (REM) sleep in the strengthening of emotional memories (e.g., Groch, Wilhelm, Diekelmann, & Born, 2012; Nishida et al., 2009; Wagner, Gais, &

Born, 2001). However, there are suggestions that this selective enhancement of emotional memories during REM sleep may also lead to enhanced susceptibility to false memories, due to a REM sleep advantage for the spread of activation of broad semantic networks (Carr & Nielsen, 2015). Investigations of sleep-dependent emotional memory are yet to thoroughly examine the different roles of valence and arousal, nor have they comprehensively investigated whether there are differences in the processing of negative versus positive emotional memories.

Chapters 4, 5, and 6 therefore aimed to disentangle effects of valence from arousal, and examined influences of both negative and positive emotion on sleep-dependent veridical and false memory performance. Previous findings of effects of emotion on memory without sleep provide mixed results, with suggestions that both emotional arousal and emotional valence may support increased memory performance. For example, Kensinger and Corkin (2003) found that negative valence and high arousal increased recognition accuracy, whereas Adelman and Estes (2013) found that both negative and positive valence supports memory consolidation, regardless of arousal. The studies presented in Chapter 4 and 6 thus help to clarify the effects of valence and arousal on memory. In Chapter 4 we found an overall bias for emotional than neutral memories, with increased veridical and false memories for both negative and positive compared to neutral words, and Chapter 6 indicated that memories rated as higher in arousal related to increased recognition accuracy. Thus, the results of this thesis suggest that selective enhancement of emotional memories occurs for information of both negative and positive valence and is influenced by the arousal rating of the memory.

Chapters 4, 5, and 6 also revealed a specific influence of sleep for emotional memory consolidation. However, in Chapter 4 sleep differentially affected the consolidation of negative and positive memories. In line with previous findings (e.g., Nishida, Pearsall, Buckner, & Walker, 2009; Payne, Stickgold, Swanberg, & Kensinger, 2008), sleep improved veridical memory for negative compared to neutral words. However, this enhanced integration of emotional memories from hippocampal to neocortical stores did not support increased false recognition. We suggest that a lack of an effect of sleep on false recognition of negative words could be explained by increased source monitoring abilities after sleep for recognition testing (Johnson et al., 1993), such that sleep supported the integration of negative emotional memories from hippocampal to neocortical systems, but was able to actively reject associated information that was activated during retrieval. Indeed, the findings of Chaper 4 do not support previous investigation of the effect of sleep on DRM veridical and false memories (McKeon, Pace-Schott, & Spencer, 2012), that found increased false recall of neutral than negative related words after sleep. The findings of McKeon et al. support a reduction in false recall of emotional information, whereas Chapter 4 suggests no difference in false recognition of emotional and neutral memories after sleep, further highlighting the differences between recall and recognition testing on emotional memory performance.

Interestingly, Chapter 4 also revealed an effect of wakefulness on positive emotion, with increased veridical and false memory for positive compared to neutral DRM words. We provide a number of potential explanations. Firstly, it may be that mood congruency effects influenced performance. There are suggestions that mood influences subsequent memory within DRM tests (Bland, Howe, & Knott, 2016; Howe & Malone, 2011; Ruci, Tomes, & Zelenski, 2009), such that higher ratings of positive mood would have related to increased false recognition of positive words. In addition, the fading affect bias (FAB; Walker, Skowronski & Thompson, 2003) and the mobilization-minimization hypothesis (Taylor, 1991) both propose that negative events are less likely to be remembered than positive events, due to cognitive and behavioural responses that aim to minimize negative emotion. Although these hypotheses are based on episodic memories, it may help to explain the findings of improved memory for positive information during wakefulness. If wakefulness does support improved memory for positive than neutral information, then it is not surprising that the results also revealed an increase in false recognition of positive words. Those in the wake group would not have benefitted from increased source monitoring abilities to the same extent as those in the sleep group (Johnson et al., 1993), and thus were less able to actively reject distantly related positive information.

Although Chapter 4 revealed potential differences in sleep-dependent memory consolidation for negative versus positive emotional memories, Chapters 5 and 6 instead support an overall bias for emotional memory after sleep, for both memory for word lists and metaphorical word pairs. In Chapter 5, we found that a short nap led to increased performance accuracy for emotional than neutral words, rather than a specific benefit for negative emotion as found in Chapter 4. Similar to Chapter 4, we found no effect of sleep on false recognition. The difference in emotional veridical memory performance between Chapter 4 and 5 may be explained by the shorter sleep time given to those in the nap study. REM sleep, which occurs in a larger proportion in the latter half of a night of sleep, has been specifically implicated in negative emotion (Nishida et al., 2009). A night of sleep including REM may therefore lead to the enhanced reactivation and integration of negative memories.

However, Chapter 6 also supports a selective enhancement of emotional memories of both negative and positive emotion after sleep compared to wake, using figurative language rather than DRM word lists. A novel finding from Chapter 6 is that emotionality led to enhanced learning and memory performance despite the meaning of the information being unknown and thus unclear in the participants' long-term memory. Using metaphorical expressions, we found that word pairs with both negative and positive meanings were remembered significantly more accurately after sleep than neutral word pairs. The nature of metaphorical expressions means that the meaning of a metaphor is not directly available, but instead needs to be learned. We found that even for novel metaphorical word pairs, in which meaning was not available from the word pair, sleep still selectively enhanced word pairs of negative and positive emotion. Chapter 6 therefore extends previous findings supporting a boost to memory for information that generates a direct emotional response (e.g., Hu et al., 2006, Wagner et al., 2001), and suggests that sleep also supports the selective enhancement of emotional memory even when the meaning of the expression is not known.

The results of Chapters 4, 5, and 6 therefore extend previous findings of sleep aiding consolidation of negative emotional information (e.g., Hu, Stylos-Allan, & Walker, 2006; Nishida et al., 2009; Payne et al., 2008), by investigating the influence of sleep-dependent memory consolidation on both negative and positive emotion compared to non-emotional information, and present some of the first findings to suggest a role of sleep in the selective enhancement of emotional metaphorical expressions.

Sleep Architecture and Emotional Memory

Chapter 5 investigated the role of sleep architecture on emotional memories, and proposed an influence of sleep spindles during slow wave sleep (SWS) on performance accuracy of emotional memories. Previous suggestions of the effects of sleep on memory indicate a specific involvement of SWS in memory consolidation and emotional memories (Marshall & Born, 2007; van der Helm & Walker, 2009; Walker, 2009; Walker & Stickgold, 2004). In particular, the information overlap to abstraction model (iOtA model; Lewis & Durrant, 2011) proposes that during SWS, reactivation of memories leads to the strengthening of these memories, supporting integration of information from hippocampal to neocortical networks. Furthermore, pharmacologically increasing sleep spindles have previously been correlated with enhanced emotional memories (Kaestner, Wixted, & Mednick, 2013), and Cairney, Durrant, Jackson, and Lewis (2014) suggest an indirect influence of sleep spindles on emotional memory accuracy. The current findings support hypotheses illustrating the involvement of SWS and sleep spindles in boosting emotional memory consolidation, and are the first to directly implicate sleep spindles during SWS to increasing emotional memory accuracy.

Sleep to Forget, Sleep to Remember Hypothesis

Chapter 6 found that after sleep word pairs were rated as significantly less arousing than after the same amount of time spent awake. The sleep to forget, sleep to remember (SFSR) hypothesis (van der Helm & Walker, 2009) predicts that sleep supports enhanced memory for emotional information, but also reduces the affective tone of the information. Although our findings support this, the effect found was quite small and did not extend to ratings of valence, thus we cannot make strong conclusions in support of the SFSR hypothesis. We therefore suggest partial support for this hypothesis and propose that we may have been more likely to see further reductions in the affective tone of information after a number of nights of sleep. Previous findings investigating the SFSR hypothesis have found no reduction in the affective tone of memories after only one night of sleep (Wagner, Fischer, & Born, 2002; Wiesner et al., 2015). Thus, although the current findings provide partial support for sleep reducing the affective tone of a memory, more research is needed before specific conclusions can be made on the influence of "sleeping to forget", and the number of sleep nights needed to show a strong effect.

7.3 Limitations

A limitation of Chapter 5 was that we conducted a nap study rather than an overnight sleep study, and thus although the use of PSG led to intriguing results as to the effects of sleep spindles on memory performance, we do not know whether similar effects would be found if participants had a full night of sleep related memory consolidation. In particular, REM sleep has been found to increase spreading activation (Cai, Mednick, Harrison, Kanady, & Mednick, 2009), and has been correlated with negative emotion (Nishida et al., 2009). Since REM sleep occurs to a greater extent in the latter half of a night of sleep, we may have found specific influences of REM sleep in an overnight sleep study. Similarly, Chapters 4 and 6 provide novel findings of the effects of sleep on both DRM memories and figurative language, however the use of actigraphy over PSG meant that we could not make conclusions as to the influences of specific sleep architecture on the selective reactivation and integration of these memories.

A second limitation is the duration of time that words were presented at recognition in Chapter 4 and 5. Words were presented to participants for 120ms before making their decision. Although this is a short presentation speed, previous DRM studies have used similar speeds (e.g., Monaghan et al., 2017). Further, previous investigation as to the effects of presentation duration on false memory rates suggests that for tests of recognition, there is no difference in performance between words presented for 20ms and those presented for 250ms (Seamon, Luo, & Gallo, 1998), although if words were presented for a longer duration of 2000ms false recognition rates did increased. Thus, we may have seen increased rates of false recognition in Chapters 4 and 5 if the presentation duration was longer. However, in both studies our word lists led to levels of false memories similar to those seen in previous DRM studies, and thus it is unlikely that presentation duration influenced our results. Similarly, all participants were exposed to the same presentation duration, thus increased veridical memory for emotional information as a consequence of sleep is unlikely to have been influenced by the short duration speed.

As already mentioned in Chapter 4, in Experiment 1 we did not gain ratings of mood from our participants. Thus, mood congruency effects may have influenced the results; increased recognition of negative words by the sleep group may be due to higher ratings of negative mood either during encoding or retrieval (Bland et al., 2016; Howe & Malone, 2011; Ruci et al., 2009). We tried to clarify this issue by collecting ratings of mood in Experiment 2 and found no difference between groups and no effect of mood on recognition accuracy, suggesting that mood is unlikely to have had an effect on performance. However, since Experiment 2 collected ratings from a different group of participants it is possible that mood may still have influenced performance in Experiment 1.

7.4 Future Directions

7.4.1 Effects of Overnight Sleep Architecture on Emotional Memory

To fully understand the effect of sleep on emotional memory processing, a future study could extend the studies presented in Chapters 4, 5, and 6 to investigate the influence of different sleep stages on emotional veridical and false memories in an overnight paradigm. As previously discussed, overnight sleep would allow us to understand the influence of REM sleep on emotional memory. Both Chapters 4 and 6 indicate a role of overnight sleep in the selective enhancement of emotional memory, however we do not know the effects of sleep architecture on this emotional memory bias. Theories of sleep and memory indicate that both SWS and REM sleep are important in the consolidation of memories; repeated reactivation of memories within the hippocampus occur during SWS and sleep spindles, and the stabilizing of these memories occurs during REM sleep (Gais et al., 2000). REM sleep in particular has been previously correlated with emotional memory (Nishida et al., 2009), however it is still unclear what effect specific sleep stages have on the consolidation of both negative and positive emotion. The use of PSG would allow us to investigate these questions.

7.4.2 Hemispheric Processing of Emotional Memories

A potential future direction would be to investigate the interaction between the hemispheric processing of memories and the influence of emotionality on memory. The findings of the thesis suggest that lateralisation effects on memory are specific to retrieval processes, and that emotional memories are selectively enhanced compared to neutral memories. However, separate literature also suggests that the processing of emotions is dependent on lateralisation effects. A RH model of emotion suggests that all emotions are processed in the RH (Borod et al., 1998), whereas the valence-arousal model suggests that the RH supports the processing of negative emotional memories and those high in arousal, and the LH supports positive emotional memories (Heller, Nitschke, & Lindsay, 1997). Findings by Mneimne, Powers, Walton, Kosson, Fonda, and Simonetti (2010) conclude that these two models work together to support emotional memory processing, with a RH bias for the processing of arousing stimuli of both negative and positive valence, and LH systems supporting the processing of positive, highly arousing information. Thus, if there is a specific enhancement for emotional memories, we should see a RH bias for both veridical and false memories high in arousal, and a LH bias for positively valenced veridical memories that are lateralised to the LH. Indeed, there are suggestions of an increase in false memories for negative stimuli that is greater in the RH (Marchewka, Brechmann, Nowicka, Jednoróg, Scheich, & Grabowska, 2008), however future studies could investigate whether there are differential effects of LH versus RH processing on negative and positive emotional memories.

7.4.3 Effect of Sleep on the Hemispheric Processing of Figurative Language

Similar to the processing of false memories, the course semantic coding hypothesis (Beeman & Bowden, 2000; Beeman & Chiarello, 1998) suggests that the RH is specifically involved in the processing of figurative language compared to literal expressions. To understand the meaning of figurative language involves the processing of distantly related concepts, and thus greater activation in the RH. This is supported by a range of studies indicating greater activation in the RH during the processing of metaphors (Anaki, Faust, & Kravetz, 1998; Bottini et al., 1994; Sotillo et al., 2004). However, many metaphors are regarded as conventional to speakers of the language, and so are familiar to them and may require less broad semantic processing. The graded salience hypothesis (Giora, 1997) predicts that there will be more LH activation for conventional figurative language, and RH activation of less salient figurative language. Future investigations could therefore examine whether a bias for false memories in the RH after sleep can be extended to other types of stimuli such as figurative language. If sleep supports accurate memory performance in the LH

(Monaghan et al., 2017), and the spread of activation of related information in the RH (Shaw & Monaghan, 2017), then we would expect sleep to support a bias in the LH for conventional metaphorical expressions, and a RH bias in memory for novel or less salient metaphorical expressions.

7.4.4 Differential Effects of Recall versus Recognition on Emotional Memory

Since previous studies of emotional false memories suggest that tests of recognition are more likely to elicit false memories than tests of recall (El Sharkawy et al., 2008; Howe et al., 2010), future research could investigate how this potential difference between recall and recognition may influence other types of emotional memory retrieval, such as during eye-witness testimony. For example, if tests of recognition are more likely to elicit false memories, this would be especially important in settings such as eye-witness testimony, where witnesses are retrieving an emotional memory and witness accuracy is necessary. Research into false recall and recognition of emotional events might help to support improvements in the way that witnesses are asked to provide critical information.

Additionally, the meta-analysis presented in Chapter 2 suggests that after sleep, tests of recall may be more likely to elicit false memories. Indeed, the studies presented in this thesis suggest no effect of sleep on false recognition, but do suggest an enhancement of veridical memory for emotional information. Thus, a benefit for recalling emotional information immediately may switch to a benefit for recognition testing after a period of offline memory consolidation. Future research could more specifically investigate the potential differences between recall and recognition testing on performance accuracy, both at immediate and delayed testing, and use various types of stimuli to provide real-world applications in areas such as eye-witness testimony.

7.5 Conclusions

In conclusion, this thesis aimed to explore the role of sleep on the processing of veridical and false memories, and outline potential factors that lead to the enhancement of memory. We contribute to the literature on DRM veridical and false memories by suggesting an effect of sleep on memory that is dependent on specific features of the task. We further extend the literature on hemispheric processing of memories to suggest that lateralisation effects are specific to retrieval processes, and illustrate that the selective enhancement of emotional memories is supported by sleep. Furthermore, we propose that sleep spindles during SWS may play a vital role in the integration of these selective emotional memories into neocortical systems. The findings of this thesis add to the growing literature on sleep and memory, and expand our understanding of the way in which sleep interacts with other factors to support memory.

References

Ackermann, S., & Rasch, B. (2014). Differential Effects of Non-REM and REM Sleep on Memory Consolidation? *Current Neurology and Neuroscience Reports, 14*, 430. doi:10.1007/s11910-013-0430-8

Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia,45*(7), 1363-1377. doi:10.1016/j.neuropsychologia.2006.10.016

Adelman, J. S., & Estes, Z. (2013). Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition, 129*(3), 530-535. doi:10.1016/j.cognition.2013.08.014

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

Anaki, D., Faust, M., & Kravetz, S. (1998). Cerebral hemispheric asymmetries in processing lexical metaphors. *Neuropsychologia, 36*(4), 353-362. doi:10.1016/S0028-3932(97)00110-3

Antony, J. W., Gobel, E. W., O'Hare, J. K., Reber, P. J., & Paller, K. A. (2012). Cued memory reactivation during sleep influences skill learning. *Nature Neuroscience, 15*(8), 1114-1116. doi:10.1038/nn.3152

Antony, J. W., Piloto, L., Wang, M., Pacheco, P., Norman, K. A., & Paller, K. A. (2018). Sleep spindle refractoriness segregates periods of memory reactivation. *Current Biology, 28*(11), 1736-1743. doi:10.1016/j.cub.2018.04.020

Antony, J. W., Schönauer, M., Staresina, B. P., & Cairney, S. A. (2018). Sleep Spindles and Memory Reprocessing. *Trends in Neurosciences.* doi:10.1016/j.tins.2018.09.012

Arndt, J., & Reder, L. M. (2003). The effect of distinctive visual information on false recognition. *Journal of Memory and Language, 48*(1), 1-15. doi:10.1016/S0749-596X(02)00518-1

Baayen, R. H. (2008). *Analyzing linguistic data [electronic resource] : a practical introduction to statistics using R*. Cambridge: Cambridge : Cambridge University Press.

Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological Influences on the Recognition of Monosyllabic Monomorphemic Words. *Journal of Memory and Language, 55*(2), 290-313. doi:10.1016/j.jml.2006.03.008

Bambini, V., Gentili, C., Ricciardi, E., Bertinetto, P. M., & Pietrini, P. (2011). Decomposing metaphor processing at the cognitive and neural level through functional magnetic resonance imaging. *Brain Research Bulletin, 86*, 203-216. doi:10.1016/j.brainresbull.2011.07.015

Barlow, J. M., Pollio, H. R., & Fine, H. J. (1977). Insight and figurative language in psychotherapy. *Psychotherapy: Theory, Research & Practice, 14*(3), 212-222. doi:10.1037/h0086530

Bauer, L. M., Olheiser, E. L., Altarriba, J., & Landi, N. (2009). Word type effects in false recall: Concrete, abstract, and emotion word critical lures. *The American Journal of Psychology, 122*(4), 469-481.

Beeman, M. J., & Bowden, E. M. (2000). The right hemisphere maintains solution-related activation for yet-to-be-solved problems. *Memory & Cognition, 28*(7), 1231-1241. doi:10.3758/BF03211823

Beeman, M. J., & Chiarello, C. (1998). Complementary right-and left-hemisphere language comprehension. *Current Directions in Psychological Science, 7*(1), 2-8. doi:10.1111/1467-8721.ep11521805

Beeman, M. J., Friedman, R. B., Grafman, J., Perez, E., Diamond, S., & Lindsay, M. B. (1994). Summation Priming and Coarse Semantic Coding in the Right Hemisphere. *Journal of Cognitive Neuroscience, 6*(1), 26-45. doi:10.1162/jocn.1994.6.1.26

Bellamy, K. J., & Shillcock, R. (2007). A right hemisphere bias towards false memory. *Laterality, 12*(2), 154-166. doi:10.1080/13576500601051648

Ben-Artzi, E., Faust, M., & Moeller, E. (2009). Hemispheric asymmetries in discourse processing: Evidence from false memories for lists and texts. *Neuropsychologia, 47*(2), 430-438. doi:10.1016/j.neuropsychologia.2008.09.021

Bland, C. E., Howe, M. L., & Knott, L. (2016). Discrete emotion-congruent false memories in the DRM paradigm. *Emotion, 16*(5), 611-619. doi:10.1037/emo0000153

Bolduc, C., Daoust, A., Limoges, É., Braun, C. M. J., & Godbout, R. (2003). Hemispheric lateralization of the EEG during wakefulness and REM sleep in young healthy adults. *Brain and Cognition, 53*(2), 193-196. doi: 10.1016/S0278-2626(03)00108-8

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*(2), 97-111. doi:10.1002/jrsm.12

Born, J., Rasch, B., & Gais, S. (2006). Sleep to Remember. *The Neuroscientist, 12*(5), 410-424. doi:10.117/1073858406292647

Borod, J. C., Cicero, B. A., Obler, L. K., Welkowitz, J., Erhan, H. M., Santschi, C., . . . Whalen, J. R. (1998). Right hemisphere emotional perception: Evidence across multiple channels. *Neuropsychology, 12*(3), 446-458. doi:10.1037/0894-4105.12.3.446

Bottini, G., Corcoran, R., Sterzi, R., Paulesu, E., Schenone, P., Scarpa, P., ... & Frith, D. (1994). The role of the right hemisphere in the interpretation of figurative aspects of language. A positron emission tomography activation study. *Brain, 117*(6), 1241-1253. doi:10.1093/brain/117.6.1241

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49-59. doi:10.1016/0005-7916(94)90063-9

Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering pictures: Pleasure and arousal in memory. *Journal of Experimental Psychology, 18*(2), 379-390. doi:10.1037/0278-7393.18.2.379

Brainerd, C. J. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science, 11*(5), 164-169. doi:10.1111/1467-8721.00192

Brainerd, C. J., Holliday, R., Reyna, V., Yang, Y., & Toglia, M. (2010). Developmental reversals in false memory: Effects of emotional valence and arousal. *Journal of Experimental Child Psychology, 107*(2), 137-154. doi:10.1016/j.jecp.2010.04.013

Brainerd, C. J., Reyna, V. F., & Ceci, S. J. (2008). Developmental reversals in false memory: A review of data and theory. *Psychological Bulletin, 134*(3), 343-382. doi:10.1037/0033-2909.134.3.343

Brainerd, C. J., Stein, L. M., Silveira, R. A., Rohenkohl, G., & Reyna, V. F. (2008). How does negative emotion cause false memories? *Psychological Science, 19*(9), 919-925. doi:10.1111/j.1467-9280.2008.02177.x

Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology, 27*(2), 307-327. doi:10.1037/0278-7393.27.2.307

Brennen, T., Dybdahl, R., & Kapidžić, A. (2007). Trauma-related and neutral false memories in war-induced Posttraumatic Stress Disorder. *Consciousness and Cognition, 16*(4), 877-885. doi:10.1016/j.concog.2006.06.012

Burgess, C., & Simpson, G. B. (1988). Cerebral hemispheric mechanisms in the retrieval of ambiguous word meanings. *Brain and Language, 33*(1), 86-103. doi:10.1016/0093-934X(88)90056-9

Buzsaki, G. (1998). Memory consolidation during sleep: a neurophysiological perspective. *Journal of Sleep Research, 7*, 17-23.

Cahill, L., Haier, R., Fallon, J., Alkire, M., Tang, C., Keator, D., . . . Mcgaugh, J. (1996). Amygdala activity at encoding correlated with long-term, free recall of emotional information. *Proceedings of the National Academy of Sciences of the United States of America, 93*(15), 8016-8021. doi:10.1073/pnas.93.15.8016

Cai, D., Mednick, S. A., Harrison, E, M., Kanady, J. C., & Mednick, S. C. (2009). REM, not incubation, improves creativity by priming associative networks. *Proceedings of the National Academy of Sciences, 106*, 10130–10134. doi:10.1073/pnas.0900271106

Cairney, S. A., Durrant, S. J., Jackson, R., & Lewis, P. A. (2014). Sleep spindles provide indirect support to the consolidation of emotional encoding contexts. *Neuropsychologia, 63,* 285 - 292. doi:10.1016/j.neuropsychologia.2014.09.016

Cardillo, E. R., Watson, C. E., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2012). From novel to familiar: Tuning the brain for metaphors. *NeuroImage, 59*(4), 3212-3221. doi:10.1016/j.neuroimage.2011.11.079

Carr, M., & Nielsen, T. (2015). Morning REM sleep naps facilitate broad access to emotional semantic networks. *Sleep, 38*(3), 433-443. doi:10.5665/sleep.4504

Carter, R. (2004). *Language and creativity [electronic resource] : The art of common talk.* London ; New York: Routledge.

Casagrande, M., & Bertini, M. (2008a). Laterality of the sleep onset process: Which hemisphere goes to sleep first? *Biological Psychology, 77*(1), 76-80. doi:10.1016/j.biopsycho.2007.09.007

Casagrande, M., & Bertini, M. (2008b). Night-time right hemisphere superiority and daytime left hemisphere superiority: A repatterning of laterality across wake–

sleep–wake states. *Biological Psychology, 77*(3), 337-342. doi:10.1016/j.biopsycho.2007.11.007

Casagrande, M., Violani, C., De Gennaro, L., Braibanti, P., & Bertini, M. (1995). Which hemisphere falls asleep first? *Neuropsychologia, 33*(7), 815-822. doi:10.1016/0028-3932(95)00004-M

Chambers, A., & Payne, J. D. (2014). Laugh yourself to sleep: memory consolidation for humorous information. *Experimental Brain Research, 232*(5), 1415-1427. doi:10.1007/s00221-013-3779-7

Chatburn, A., Lushington, K., & Kohler, M. J. (2014). Complex associative memory processing and sleep: A systematic review and meta-analysis of behavioural evidence and underlying EEG mechanisms. *Neuroscience and Biobehavioral Reviews, 47*, 646-655. doi:10.1016/j.neubiorev.2014.10.018

Chiarello, C., Burgess, C., Richards, L., & Pollock, A. (1990). Semantic and associative priming in the cerebral hemispheres: Some words do, some words don't … sometimes, some places. *Brain and Language, 38*(1), 75-104. doi:10.1016/0093-934X(90)90103-N

Citron, F. M. M., & Goldberg, A. E. (2014). Metaphorical sentences are more emotionally engaging than their literal counterparts. *Journal of Cognitive Neuroscience, 26*(11), 2585-2595. doi:10.1162/jocn_a_00654

Citron, F. M. M., Güsten, F., Michaelis, N., & Goldberg, A. E. (2016). Conventional metaphors in longer passages evoke affective brain response. *NeuroImage, 139,* 218-230. doi:10.1016/j.neuroimage.2016.06.020

Clemens, Fabó, & Halász. (2006). Twenty-four hours retention of visuospatial memory correlates with the number of parietal sleep spindles. *Neuroscience Letters, 403*(1), 52-56. doi:10.1016/j.neulet.2006.04.035

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407-428. doi:10.1037/0033-295X.82.6.407

Cox, R., Hofman, W. F., & Talamini, L. M. (2012). Involvement of spindles in memory consolidation is slow wave sleep-specific. *Learning & Memory, 19*(7), 264-267. doi:10.1101/lm.026252.112

Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience, 21*(4), 803-820. doi:10.1162/jocn.2009.21059

Dennis, N. A., Kim, H., & Cabeza, R. (2007). Effects of aging on true and false memory formation: An fMRI Study. *Neuropsychologia, 45*(14), 3157-3166. doi:10.1016/j.neuropsychologia.2007.07.003

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience, 11*(2), 114-126. doi:10.1038/nrn2762

Diekelmann, S., Born, J., & Wagner, U. (2010). Sleep enhances false memories depending on general memory performance. *Behavioural Brain Research, 208*(2), 425-429. doi:10.1016/j.bbr.2009.12.021

Diekelmann, S., Büchel, C., Born, J., & Rasch, B. (2011). Labile or stable: Opposing consequences for memory when reactivated during waking and sleep. *Nature Neuroscience, 14*(3), 381-386. doi:10.1038/nn.2744

Diekelmann, S., Landolt, H., Lahl, O., Born, J., & Wagner, U. (2008). Sleep loss produces false memories. *PloS one, 3*(10), e3512-e3512. doi:10.1371/journal.pone.0003512

Dienes, Z., & McLatchie, N. (2017). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin & Review, 25*(1), 207-218. doi:10.3758/s13423-017-1266-z.

Dumay, N., & Gaskell, G. (2007). Sleep-Associated Changes in the Mental Representation of Spoken Words. *Psychological Science, 18*(1), 35-39. doi:10.1111/j.1467-9280.2007.01845.x

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629-634.

Eichenbaum, H. (2000). A cortical-hippocampal system for declarative memory. *Nature Reviews Neuroscience, 1*(1), 41-50. doi:10.1038/35036213

El Sharkawy, J., Groth, K., Vetter, C., Beraldi, A., & Fast, K. (2008). False memories of emotional and neutral words. *Behavioural Neurology, 19*, 7-11. doi:10.1155/2008/587239

Ellis, A. (2004). Length, formats, neighbours, hemispheres, and the processing of words presented laterally or at fixation. *Brain and Language, 88*(3), 355-366. doi:10.1016/S0093-934X(03)00166-4

Fabiani, M., Stadler, M. A., & Wessels, P. M. (2000). True but not false memories produce a sensory signature in human lateralized brain potentials. *Journal of Cognitive Neuroscience, 12*(6), 941-949. doi:10.1162/08989290051137486

Fainsilber, L., & Ortony. A. (1987). Metaphorical uses of language in the expression of emotions. *Metaphor and Symbolic Activity, 2*(4), 239-250. doi:10.1207/s15327868ms0204_2

Faust, M., Ben-Artzi, E., & Harel, I. (2008). Hemispheric asymmetries in semantic processing: Evidence from false memories for ambiguous words. *Brain and Language, 105*(3), 220-228. doi:10.1016/j.bandl.2007.12.002

Fenn, K. M., Gallo, D. A., Margoliash, D., Roediger, H. L., & Nusbaum, H. C. (2009). Reduced false memory after sleep. *Learning & Memory, 16*(9), 509-513. doi:10.1101/lm.1500808

Ferrarelli, F., Huber, R., Peterson, M. J., Massimini, M., Murphy, M., Riedner, B. A., Watson, A., Bria, P., Tononi, G. (2007). Reduced sleep spindle activity in schizophrenia patients. *American Journal of Psychiatry, 164*, 483–492. doi:10.1176/appi.ajp.164.3.483

Fogel, S. M., & Smith, C. T. (2006). Learning-dependent changes in sleep spindles and Stage 2 sleep. *Journal of Sleep Research, 15*(3), 250-255. doi:10.1111/j.1365-2869.2006.00522.x

Forgács, B., Lukács, Á., & Pléh, C. (2014). Lateralized processing of novel metaphors: Disentangling figurativeness and novelty. *Neuropsychologia, 56*, 101-109. doi:10.1016/j.neuropsychologia.2014.01.003

Fowler, M. J., Sullivan, M. J., & Ekstrand, B. R. (1973). Sleep and memory. *Science, 179*(4070), 302-304.

Frankland, P.W. & Bontempi, B. (2005). The organization of recent and remote memories. *Nature Reviews Neuroscience, 6*(2), 119-130. doi: 10.1038/nrn1607

Frost, R. L. A., & Monaghan, P. (2017). Sleep-driven computations in speech processing. *PLoS One, 12*(1): e0169538. doi:10.1371/journal.pone.0169538.

Gais, S., & Born, J. (2004). Declarative Memory Consolidation: Mechanisms Acting during Human Sleep. *Learning & Memory, 11*(6), 679-685. doi:10.1101/lm.80504

Gais, S., & Born, J. (2004). Low Acetylcholine during Slow-Wave Sleep Is Critical for Declarative Memory Consolidation. *Proceedings of the National Academy of Sciences of the United States of America,101*(7), 2140-2144. doi:10.1073/pnas.0305404101

Gais, S., Mölle, M., Helms, K., & Born, J. (2002). Learning-dependent increases in sleep spindle density. *The Journal of Neuroscience, 22*(15), 6830-6834. doi:10.1523/JNEUROSCI.22-15-06830.2002

Gais, S., Plihal, W., Wagner, U., & Born, J. (2000). Early sleep triggers memory for early visual discrimination skills. *Nature Neuroscience, 3*(12), 1335-9. doi:10.1038/81881

Gallo, D. A., & Roediger, H. L. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language, 47*(3), 469-497. doi:10.1016/S0749-596X(02)00013-X

Gallo, D. A., & Roediger, H. L. (2003). The effects of associations and aging on illusory recollection. *Memory & Cognition, 31*(7), 1036-1044. doi:10.3758/BF03196124

Gibbons, J. A., Lee, S. A., & Walker, W. R. (2011). The fading affect bias begins within 12 hours and persists for 3 months. *Applied Cognitive Psychology, 25*(4), 663-672. doi:10.1002/acp.1738

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics, 8*(3), 183-206.

Giora, R. (1999). On the priority of salient meanings: Studies of literal and figurative language. *Journal of Pragmatics, 31*(7), 919-929. doi:10.1016/S0378-2166(98)00100-3

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

Glucksberg, S. (1998). Understanding Metaphors. *Current Directions in Psychological Science, 7*(2), 39-43. 10.1111/1467-8721.ep13175582

Goldstein, A. N., & Walker, M. P. (2014). The role of sleep in emotional brain function. *Annual Review of Clinical Psychology, 10*, 679-708. doi:10.1146/annurev-clinpsy-032813-153716

Groch, S., Wilhelm, I., Diekelmann, S., & Born, J. (2012). The role of REM sleep in the processing of emotional memories: Evidence from behavior and event-related potentials. *Neurobiology of Learning and Memory, 99*, 1-9. doi:10.1016/j.nlm.2012.10.006

Heller, W., Nitschke, J. B., & Lindsay, D. L. (1997) Neuropsychological correlates of arousal in self-reported emotion. *Cognition and Emotion, 11*(4), 383-402. doi:10.1080/026999397379854

Hocking, J., McMahon, K. L., & De Zubicaray, G. I. (2009). Semantic context and visual feature effects in object naming: An fMRI study using arterial spin labeling. *Journal of Cognitive Neuroscience,21*(8), 1571-1583. doi:10.1162/jocn.2009.21114

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. (1973). Quantification of Sleepiness: A New Approach. *Psychophysiology, 10*(4), 431-436.

Hoddes, R., Dement, W., & Zarcone, V. (1972). The development and use of the Stanford Sleepiness Scale (SSS). *Psychophysiology, 9*, 150.

Howe, M. L., & Malone, C. (2011). Mood-congruent true and false memory: Effects of depression. *Memory, 19*(2), 192-201. doi:10.1080/09658211.2010.544073

Howe, M. L., & Wilkinson, S. (2011). Using story contexts to bias children's true and false memories. *Journal of Experimental Child Psychology, 108*(1), 77-95. doi:10.1016/j.jecp.2010.06.009

Howe, M. L., Candel, I., Otgaar, H., Malone, C., & Wimmer, M. C. (2010). Valence and the development of immediate and long-term false memory illusions. *Memory, 18*(1), 58-75. doi:10.1080/09658210903476514

Howe, M. L., Wimmer, M. C., Gagnon, N., & Plumpton, S. (2009). An associative-activation theory of children's and adults' memory illusions. *Journal of Memory and Language, 60*(2), 229-251. doi:10.1016/j.jml.2008.10.002

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

Ito, T., Cacioppo, J., & Lang, P. (1998). Eliciting Affect Using the International Affective Picture System: Trajectories through Evaluative Space. *Personality and Social Psychology Bulletin, 24*(8), 855-879. doi:10.1177/0146167298248006

Ito, Y. (2001). Hemispheric asymmetry in the induction of false memories. *Laterality, 6*(4), 337-346. doi:10.1080/13576500143000050

Jackendoff, R. (1995). *'The boundaries of the lexicon'* in M. Everaert, E. van der Linden, A. Schenk, and R. Schreuder (eds): Idioms: Structural and Psychological Perspectives. Hillsdale, NJ: Lawrence Erlbaum, pp. 133–66.

Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (Transformation or Not) and towards Logit Mixed Models. *Journal of Memory and Language, 59*(4), 434-446. doi:10.1016/j.jml.2007.11.007

Jeffreys, H. (1939/1961). *The theory of probability (1st/3rd Ed.).* Oxford: Oxford University Press.

Jelinek, L., Hottenrott, B., Randjbar, S., Peters, M. J., & Moritz, S. (2009). Visual false memories in post-traumatic stress disorder (PTSD). *Journal of Behavior Therapy and Experimental Psychiatry, 40*(2), 374-383. doi:10.1016/j.jbtep.2009.02.003

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

Johnson, M., Hashtroudi, S., Lindsay, D., & Steinberg, Robert J. (1993). Source Monitoring. *Psychological Bulletin, 114*(1), 3-28. doi:10.1037/0033-2909.114.1.3

Kaestner, E. J., Wixted, J. T., & Mednick, S. C. (2013). Pharmacologically increasing sleep spindles enhances recognition for negative and high- arousal memories. *Journal of Cognitive Neuroscience, 25*(10), 1597. doi:10.1162/jocn_a_00433

Kattler, H., Dijk, D., & Borbely, A. A. (1994). Effect of unilateral somatosensory stimulation prior to sleep on the sleep EEG in humans. *Journal of Sleep Research*, *3*, 159-164. doi:10.1111/j.1365-2869.1994.tb00123.x

Kellogg, R. T. (2001). Presentation modality and mode of recall in verbal false memory. *Journal of Experimental Psychology, 27*(4), 913-919. doi:10.1037/0278-7393.27.4.913

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition, 31*(8), 1169-1180. doi:10.3758/BF03195800

Kensinger, E. A., & Corkin, S. (2004a). Two routes to emotional memory: distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences of the United States of America, 101*(9), 3310-3315. doi:10.1073/pnas.0306408101

Kensinger, E. A., & Corkin, S. (2004b). The effects of emotional content and aging on false memories. *Cognitive, Affective, & Behavioral Neuroscience, 4*(1), 1-9. doi:10.3758/CABN.4.1.1

Kousta, S. T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition, 112*(3), 473-481. doi:10.1016/j.cognition.2009.06.007

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. *Trends in Cognitive Sciences, 20*(7), 512-534. doi:10.1016/j.tics.2016.05.004

Labar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience, 7*(1), 54-64. doi:10.1038/nrn1825

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

Lampinen, J. M., & Schwartz, R. M. (2000). The impersistence of false memory persistence. *Memory, 8*(6), 393-400. doi:10.1080/09658210050156840

Lang, A., Dhillon, K., & Dong, Q. (1995). The effects of emotional arousal and valence on television viewers' cognitive capacity and memory. *Journal of Broadcasting & Electronic Media, 39*(3), 313-327. doi:10.1080/08838159509364309

Lechner, H., Squire, L., & Byrne, J. (1999). 100 years of consolidation--remembering Müller and Pilzecker. *Learning & Memory, 6*(2), 77-87.

Levine, B., Svoboda, E., Hay, J. F., Winocur, G., & Moscovitch, M. (2002). Aging and autobiographical memory: Dissociating episodic from semantic retrieval. *Psychology and Aging, 17*(4), 677-689.

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

Liu, P. Q., Connell, L., & Lynott, D. (2018). Ease-of-Simulation Norms for 452 Adjective Metaphors. *Manuscript under Review*, 1–27.

Lo, J. C., Sim, S. K. Y., & Chee, M. W. L. (2014). Sleep reduces false memory in healthy older adults. *Sleep, 37*(4), 665-671. doi:10.5665/sleep.3564

Marchewka, A., Brechmann, A., Nowicka, A., Jednoróg, K., Scheich, H., & Grabowska, A. (2008). False recognition of emotional stimuli is lateralised in the brain: An fMRI study. *Neurobiology of Learning and Memory ,90*(1), 280-284. doi:10.1016/j.nlm.2008.01.012

Marr, D. (1971). Simple Memory: A Theory for Archicortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences (1934-1990), 262*(841), 23-81. doi:10.1098/rstb.1971.0078

Marshall, L., & Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends in Cognitive Sciences, 11*(10), 442-450. doi:10.1016/j.tics.2007.09.001

Marshall, L., Helgadóttir, H., Mölle, M., & Born, J. (2006). Boosting slow oscillations during sleep potentiates memory. *Nature, 444*(7119), 610-613. doi:10.1038/nature05278

Mashal, N., & Itkes, O. (2013). The effects of emotional valence on hemispheric processing of metaphoric word pairs. *Laterality: Asymmetries of Body, Brain and Cognition*, *19*(5), 511-521. doi: 10.1080/1357650X.2013.862539

Mather, M., & Sutherland, M. (2009). Disentangling the Effects of Arousal and Valence on Memory for Intrinsic Details. *Emotion Review, 1*(2), 118-119. doi:10.1177/1754073908100435

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*(3), 419–457. doi:10.1037/0033-295X.102.3.419

McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language, 35*(2), 212-230. doi:10.1006/jmla.1996.0012

McDermott, K. B., & Watson, J. M. (2001). The rise and fall of false recall: The impact of presentation duration. *Journal of Memory and Language, 45*(1), 160-176. doi:10.1006/jmla.2000.2771

McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience, 27*(1), 1-28. doi:10.1146/annurev.neuro.27.070203.144157

McKeon, S., Pace-Schott, E. F., & Spencer, R. M. C. (2012). Interaction of sleep and emotional content on the production of false memories. *PloS one, 7*(11), e49353. doi:10.1371/journal.pone.0049353

Mednick, S., Nakayama, K., & Stickgold, R. (2003). Sleep-dependent learning: A nap is as good as a night. *Nature Neuroscience, 6*(7), 697-698. doi:10.1038/nn1078

Mirković, J., & Gaskell, M. G. (2016). Does sleep improve your grammar? Preferential consolidation of arbitrary components of new linguistic knowledge. *PloS one, 11*(4), e0152489. doi:10.1371/journal.pone.0152489

Mneimne, M., Powers, A. S., Walton, K. E., Kosson, D. S., Fonda, S., & Simonetti, J. (2010). Emotional valence and arousal effects on memory and hemispheric asymmetries. *Brain and Cognition, 74*(1), 10-17. doi:10.1016/j.bandc.2010.05.011

Mölle, M., Marshall, L., Gais, S., & Born, J. (2002). Grouping of spindle activity during slow oscillations in human non-rapid eye movement sleep. *The Journal of Neuroscience, 22*(24), 10941-10947. doi:10.1523/JNEUROSCI.22-24-10941.2002

Monaghan, P., Shaw, J. J., Ashworth-Lord, A., & Newbury, C. R. (2017). Hemispheric processing of memory is affected by sleep. *Brain and Language, 167*, 36-43. doi:10.1016/j.bandl.2016.05.003

Monaghan, P., Shillcock, R., & McDonald, S. (2004). Hemispheric asymmetries in the split-fovea model of semantic processing. *Brain and Language, 88*(3), 339-354. doi:10.1016/S0093-934X(03)00165-2

Moscovitch, M., Rosenbaum, R., Gilboa, A., Addis, D., Westmacott, R., Grady, C., . . . Nadel, L. (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: A unified account based on multiple trace theory. *Journal of Anatomy, 207*(1), 35-66. doi:10.1111/j.1469-7580.2005.00421.x

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology, 7*(2), 217-227. doi:10.1016/S0959-4388(97)80010-4

Nadel, L., Samsonovich, A., Ryan, L., & Moscovitch, M. (2000). Multiple trace theory of human memory: Computational, neuroimaging, and neuropsychological results. *Hippocampus, 10*(4), 352-368. doi:10.1002/1098-1063(2000)10:4<352::AID-HIPO2>3.0.CO2-D

Nelson, D., McEvoy, C., & Schreiber, T. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 402-407. doi:10.3758/BF03195588

Neuschatz, J. S., Payne, D. G., Lampinen, J. M., & Toglia, M. P. (2001). Assessing the effectiveness of warnings and the phenomenological characteristics of false memories. *Memory, 9*(1), 53-71. doi:10.1080/09658210042000076

Newbury, C. R., & Monaghan, P. (2018). *Negative but not Positive Emotional Memories are Enhanced by Sleep.* Manuscript submitted for publication.

Newbury, C. R., & Monaghan, P. (2018). When does sleep affect veridical and false memory consolidation?: A meta-analysis. *Psychonomic Bulletin & Review*, 1-14.

Nishida, M., Pearsall, J., Buckner, R. L., & Walker, M. P. (2009). REM sleep, prefrontal theta, and the consolidation of human emotional memory. *Cerebral Cortex, 19*(5), 1158-1166. doi:10.1093/cercor/bhn155

O'Reilly, R. C, & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework. *Trends in Cognitive Sciences, 6*(12), 505-510. doi:10.1016/S1364-6613(02)02005-3

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113.

Pardilla-Delgado, & Payne. (2017). The impact of sleep on true and false memory across long delays. *Neurobiology of Learning and Memory, 137*, 123-133. doi:10.1016/j.nlm.2016.11.016

Pavlides, C., & Winson, J. (1989). Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. *The Journal of Neuroscience, 9*(8), 2907-2918. doi:10.1523/JNEUROSCI.09-08-02907.1989

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory Illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language, 35*(2), 261-285. doi:10.1006/jmla.1996.0015

Payne, J. D., Schacter, D. L., Propper, R. E., Huang, L., Wamsley, E. J., Tucker, M. A., . . . Stickgold, R. (2009). The role of sleep in false memory formation. *Neurobiology of Learning and Memory, 92*(3), 327-334. doi:10.1016/j.nlm.2009.03.007

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

Pesta, B. J., Murphy, M. D., & Sanders, R. E. (2001). Are emotionally charged lures immune to false memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(2), 328-338. doi:10.1037/0278-7393.27.2.328

Peters, K. R., Ray, L., Smith, V., & Smith. (2008). Changes in the density of stage 2 sleep spindles following motor learning in young and older adults. *Journal of Sleep Research, 17*(1), 23-33. doi: 10.1111/j.1365-2869.2008.00634.x

Pirozzolo, F. J., & Rayner, K. (1977). Hemispheric specialization in reading and word recognition. *Brain and Language, 4*(2), 248-261. doi:10.1016/0093-934X(77)90021-9

Plihal, W., & Born, J. (1997). Effects of early and late nocturnal sleep on declarative and procedural memory. *Journal of Cognitive Neuroscience, 9*(2), 534-547. doi:10.1162/jocn.1997.9.4.534

Pynte, J., Besson, M., Robichon, F., & Poli, J. (1996). The Time-Course of Metaphor Comprehension: An Event-Related Potential Study. *Brain and Language, 55*(3), 293-316. doi:10.1006/brln.1996.0107

Qin, Y. L., McNaughton, B. L., Skaggs, W. E., & Barnes, C. A. (1997). Memory reprocessing in corticocortical and hippocampocortical neuronal ensembles. *Philosophical Transactions, 352*(1360), 1525-1533. doi:10.1098/rstb.1997.0139

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rasch, B., & Born, J. (2007). Maintaining memories by reactivation. *Current Opinion in Neurobiology, 17*(6), 698-703. doi:10.1016/j.conb.2007.11.007

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

Rasch, B., Büchel, C., Gais, S., & Born, J. (2007). Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science, 315*(5817), 1426-14269. doi:10.1126/science.1138581

Reder, L. M., Park, H., & Kieffaber, P. D. (2009). Memory systems do not divide on consciousness: Reinterpreting memory in terms of activation and binding. *Psychological Bulletin, 135*(1), 23-49. doi:10.1037/a0013974

Reisenzein, R. (1994). Pleasure-Arousal Theory and the intensity of emotions. *Journal of Personality and Social Psychology, 67*(3), 525-539. doi:10.1037/0022-3514.67.3.525

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7*(1), 1-75. doi:10.1016/1041-6080(95)90031-4

Reyna, V. F., Corbin, J. C., Weldon, R. B., & Brainerd, C. J. (2016). How fuzzy-trace theory predicts true and false memories for words, sentences, and narratives. *Journal of Applied Research in Memory and Cognition, 5*(1), 1-9. doi:10.1016/j.jarmac.2015.12.003

Ribot, T. (1882). *Diseases of memory*. New York: Appleton-Century-Crofts.

Riedel, G., & Micheau, J. (2001). Function of the hippocampus in memory formation: Desperately seeking resolution. *Progress in Neuropsychopharmacology & Biological Psychiatry, 25*(4), 835-853. doi:10.1016/S0278-5846(01)00153-1

Ritchie, T. D., Batteson, T. J., Bohn, A., Crawford, M. T., Ferguson, G. V., Schrauf, R. W., . . . Walker, W. R. (2015). A pancultural perspective on the fading affect bias in autobiographical memory. *Memory, 23*(2), 278-290. doi:10.1080/09658211.2014.884138

Robinson, K. J., & Roediger, H. L. (1997). Associative processes in false recall and false recognition. *Psychological Science, 8*(3), 231-237. doi:10.1111/j.1467-9280.1997.tb00417.x

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology, 21*(4), 803-814. doi:10.1037/0278-7393.21.4.803

Roediger, H. L., & McDermott, K. B. (2000). Tricks of memory. *Current Directions in Psychological Science, 9*(4), 123-127. doi:10.1111/1467-8721.00075

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8*(3), 385-407. doi:10.3758/BF03196177

Rosanova, M., & Ulrich, D. (2005). Pattern-specific associative long-term potentiation induced by a sleep spindle-related spike train. *The Journal of Neuroscience, 25*(41), 9398-9405. doi:10.1523/JNEUROSCI.2149-05.2005

Ruci, L., Tomes, J. L., & Zelenski, J. M. (2009). Mood-congruent false memories in the DRM paradigm. *Cognition & Emotion, 23*(6), 1153-1165. doi:10.1080/02699930802355420

Schacter, D. L., & Slotnick, S. D. (2004). The cognitive neuroscience of memory distortion. *Neuron, 44*(1), 149-160. doi:10.1016/j.neuron.2004.08.017

Schmidt, C., Collette, F., Cajochen, C., & Peigneux, P. (2007). A time to think: Circadian rhythms in human cognition. *Cognitive Neuropsychology, 24*(7), 755-789. doi:10.1080/02643290701754158

Schmidt, C., Peigneux, P., Muto, V., Schenkel, M., Knoblauch, V., Münch, M., . . . Cajochen, C. (2006). Encoding difficulty promotes postlearning changes in sleep spindle activity during napping. *The Journal of Neuroscience, 26*(35), 8976-8982. doi:10.1523/JNEUROSCI.2464-06.2006

Schmidt, G. L., & Seger, C. A. (2009). Neural correlates of metaphor processing: The roles of figurativeness, familiarity and difficulty. *Brain and Cognition, 71*(3), 375-386. doi:10.1016/j.bandc.2009.06.001

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry, 20*(1), 11–21. doi:10.1136/jnnp.20.1.11

Seamon, J. G., Luo, C. R., & Gallo, D. A. (1998). Creating false memories of words with or without recognition of list items: Evidence for nonconscious processes. *Psychological Science, 9*(1), 20-26. doi:10.1111/1467-9280.00004

Sharkawy, J. E., Groth, K., Vetter, C., Beraldi, A., & Fast, K. (2008). False memories of emotional and neutral words. *Behavioural Neurology, 19*(1), 7-11. doi:10.1155/2008/587239

Shaw, J. J., & Monaghan, P. (2017). Lateralised sleep spindles relate to false memory generation. *Neuropsychologia, 107*, 60-67. doi:10.1016/j.neuropsychologia.2017.11.002

Siapas, A. G., & Wilson, M. A. (1998). Coordinated interactions between hippocampal ripples and cortical spindles during slow-wave sleep. *Neuron, 21*(5), 1123-1128. doi:10.1016/S0896-6273(00)80629-7

Sio, U., Monaghan, P., & Ormerod, T. (2013). Sleep on it, but only if it is difficult: Effects of sleep on problem solving. *Memory & Cognition, 41*(2), 159-166. doi:10.3758/s13421-012-0256-7

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 of the United States of America, 100*(4), 2065-2069. doi:10.1073/pnas.0437938100

Smith, R. E., & Engle, R. W. (2011). Study modality and false recall. *Experimental Psychology, 58*(2), 117-124. doi:10.1027/1618-3169/a000076

Smith, R. E., & Hunt, R. R. (1998). Presentation Modality Affects False Memory. *Psychonomic Bulletin & Review, 5*(4), 710-715. doi:10.3758/BF03208850

Smith, T. A., & Kimball, D. R. (2012). Revisiting the rise and fall of false recall: Presentation rate effects depend on retention interval. *Memory, 20*(6), 535-553. doi:10.1080/09658211.2012.684881

Sotillo, M., Carretié, L., Hinojosa, J. A., Tapia, M., Mercado, F., López-Martín, S., & Albert, J. (2004). Neural activity associated with metaphor comprehension: spatial analysis. *Neuroscience Letters, 373*(1), 5-9. doi:10.1016/j.neulet.2004.09.071

Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences of the United States, 93*(24), 13515-13522. doi:10.1073/pnas.93.24.13515

Squire, L. R., Cohen, N. J., & Nadel, L. (1984). *The medial temporal region and memory consolidation: A new hypothesis.* In H.Weingartner & E.Parker (Eds.), Memory consolidation (pp. 185–210). Hillsdale, NJ: Erlbaum.

Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition, 27*(3), 494-500. doi:10.3758/BF03211543

Stickgold, & Walker. (2005). Memory consolidation and reconsolidation: What is the role of sleep? *Trends in Neurosciences, 28*(8), 408-415. doi:10.1016/j.tins.2005.06.004

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature Cell Biology, 437*(7063), 1272-1278. doi:10.1038/nature04286

Stickgold, R., & Walker, M. (2004). To sleep, perchance to gain creative insight? *Trends in Cognitive Sciences, 8*(5), 191-192. doi:10.1016/j.tics.2004.03.003

Straube, B. (2012). An overview of the neuro-cognitive processes involved in the encoding, consolidation, and retrieval of true and false memories. *Behavioral and Brain Functions, 8*(1), 35. doi: 10.1186/1744-9081-8-35

Takashima, A., Petersson, K. M., Rutters, F., Tendolkar, I., Jensen, O., Zwarts, M. J., ... & Fernandez, G. (2006). Declarative memory consolidation in humans: a prospective functional magnetic resonance imaging study. *Proceedings of the National Academy of Sciences*, *103*(3), 756-761. doi:10.1073 pnas.0507774103

Tamminen, J., Lambon Ralph, M., & Lewis, P. (2013). The role of sleep spindles and slow-wave activity in integrating new information in semantic memory. *The Journal of Neuroscience, 33*(39), 15376-81. doi:10.1523/JNEUROSCI.5093-12.2013

Tamminen, J., Payne, J., Stickgold, R., Wamsley, E., & Gaskell, M. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *The Journal of Neuroscience, 30*(43), 14356-14360. doi:10.1523/JNEUROSCI.3028-10.2010

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: the mobilization-minimization hypothesis. *Psychological Bulletin, 110*(1), 67-85. doi:10.1037/0033-2909.110.1.67

Thapar, A., & McDermott, K. (2001). False recall and false recognition induced by presentation of associated words: Effects of retention interval and level of processing. *Memory & Cognition, 29*(3), 424-432. doi:10.3758/BF03196393

Toglia, M., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory, 7*(2), 233-256. doi:10.1080/741944069

Tucker, M., Hirota, Y., Wamsley, E., Lau, H., Chaklader, A., & Fishbein, W. (2006). A daytime nap containing solely non-REM sleep enhances declarative but not procedural memory. *Neurobiology of Learning and Memory, 86*(2), 241-247. doi:10.1016/j.nlm.2006.03.005

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*(1), 1-12. doi:10.1037/h0080017

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

van der Helm, E., Yao, J., Dutt, S., Rao, V., Saletin, J., & Walker, M. (2011). REM sleep depotentiates amygdala activity to previous emotional experiences. *Current Biology, 21*(23), 2029-2032. doi:10.1016/j.cub.2011.10.052

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67*(6), 1176-1190. doi:10.1080/17470218.2013.850521

Vazquez, J., & Baghdoyan, H. A. (2001). Basal forebrain acetylcholine release during REM sleep is significantly greater than during waking. *American Journal of Physiology, 280*(2), R598-R601. doi:10.1152/ajpregu.2001.280.2.R598

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software, 36*(3). doi:10.18637/jss.v036.i03

Violani, C., Testa, P., & Casagrande, M. (1998). Actigraphic motor asymmetries during sleep. *Sleep, 21*(5), 472-476.

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*(2), 112-9. doi:10.1101/lm.36801

Wagner, U., Hallschmid, M., Verleger, R., & Born, J. (2003). Signs of REM sleep dependent enhancement of implicit face memory: A repetition priming study. *Biological Psychology, 62*(3), 197-210. doi:10.1016/S0301-0511(02)00125-4

Wagner, U., Fischer, S., & Born, J. (2002). Changes in emotional responses to aversive pictures across periods rich in slow-wave sleep versus rapid eye movement sleep. *Psychosomatic Medicine, 64*(4), 627-634.

Walker, M. P. (2009). The Role of Sleep in Cognition and Emotion. *Annals of the New York Academy of Sciences, 11561*(1), 168-197. doi:10.1111/j.1749-6632.2009.04416.x

Walker, M. P., Brakefield, T., Morgan, A., Hobson, J., & Stickgold, R. (2002). Practice with Sleep Makes Perfect: Sleep-Dependent Motor Skill Learning. *Neuron, 35*(1), 205-211. doi:10.1016/S0896-6273(02)00746-8

Walker, W. P., Skowronski, J. J., & Thompson, C. P. (2003). Life is pleasant—and memory helps to keep it that way! *Review of General Psychology, 7*(2), 203-210. doi:10.1037/1089-2680.7.2.203

Walker, M. P., & Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron, 44*(1), 121-133. doi:10.1016/j.neuron.2004.08.031

Walker, M. P., & Stickgold, R. (2006). Sleep, memory, and plasticity. *Annual Review of Psychology, 57*, 139-166. doi:10.1146/annurev.psych.56.091103.070307

Walker, W. P., Vogl, R. J., & Thompson, C. P. (1997). Autobiographical memory: unpleasantness fades faster than pleasantness over time. *Applied Cognitive Psychology, 11*(5), 399-413. doi:10.1002/(SICI)1099-0720(199710)11:5<399::AID-ACP462>3.0.CO;2-E

Warriner, A., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(4), 1191-1207. doi:10.3758/s13428-012-0314-x

Watson, J. M., McDermott, K. B., & Balota, D. A. (2004). Attempting to avoid false memories in the Deese/Roediger-McDermott paradigm: Assessing the combined influence of practice and warnings in young and old adults. *Memory & Cognition, 32*(1), 135-141. doi:10.3758/BF03195826.

Westerberg, C. E., & Marsolek, C. J. (2003). Hemisphere asymmetries in memory processes as measured in a false recognition paradigm. *Cortex, 39*(4-5), 627-642. doi:10.1016/S0010-9452(08)70857-8

Whitehead, R. (1991). Right hemisphere processing superiority during sustained visual attention. *Journal of Cognitive Neuroscience, 3*(4), 329-334. doi:10.1162/jocn.1991.3.4.329

Wiesner, C. D., Pulst, J., Krause, F., Elsner, M., Baving, L., Pedersen, A., . . . Göder, R. (2015). The effect of selective REM-sleep deprivation on the consolidation and affective evaluation of emotional memories. *Neurobiology of Learning and Memory, 122,* 131-141. doi:10.1016/j.nlm.2015.02.008

Wilson, J. K., Baran, B., Pace-Schott, E. F., Ivry, R. B., & Spencer, R. M. C. (2012). Sleep modulates word-pair learning but not motor sequence learning in healthy older adults. *Neurobiology of Aging, 33*(5), 991-1000. doi:10.1016/j.neurobiolaging.2011.06.029

Wilson, J., Baran, B., Pace-Schott, E. F., Ivry, R. B., & Spencer, R. M. C. (2012). Sleep modulates word-pair learning but not motor sequence learning in healthy older adults. *Neurobiology of Aging, 33*(5), 991-1000. doi:10.1016/j.neurobiolaging.2011.06.029

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

Winocur, G., & Moscovitch, M. (2011). Memory transformation and systems consolidation. *Journal of the International Neuropsychological Society, 17*(5), 766-780. doi: 10.1017/S1355617711000683

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

Zoellner, L. A., Foa, E. B., Brigidi, B. D., & Przeworski, A. (2000). Are trauma victims susceptible to "false memories"? *Journal of Abnormal Psychology, 109*(3), 517-524. doi:10.1037//0021-843x.109.3.517

Appendices

Appendix A. DRM word lists used in Chapter 3, taken from Stadler et al., (1999).

BREAD: butter, food, eat, sandwich, rye, jam, milk, flour, jelly, dough, crust, slice, wine, loaf, toast

CAR: truck, bus, train, automobile, vehicle, drive, jeep, Ford, race, keys, garage, highway, sedan, van, taxi

CHAIR: table, sit, legs, seat, couch, desk, recliner, sofa, wood, cushion, swivel, stool, sitting, rocking, bench

DOCTOR: nurse, sick, lawyer, medicine, health, hospital, dentist, physician, ill, patient, office, stethoscope, surgeon, clinic, cure

FRUIT: apple, vegetable, orange, kiwi, citrus, ripe, pear, banana, berry, cherry, basket, juice, salad, bowl, cocktail

KING: queen, England, crown, prince, George, dictator, palace, throne, chess, rule, subjects, monarch, royal, leader, reign

MOUNTAIN: hill, valley, climb, summit, top, molehill, peak, plain, glacier, goat, bike, climber, range, steep, ski

MUSIC: note, sound, piano, sing, radio, band, melody, horn, concert, instrument, symphony, jazz, orchestra, art, rhythm

NEEDLE: thread, pin, eye, sewing, sharp, point, prick, thimble, haystack, thorn, hurt, injection, syringe, cloth, knitting

RIVER: water, stream, lake, Mississippi, boat, tide, swim, flow, run, barge, creek, brook, fish, bridge, winding

SLEEP: bed, rest, awake, tired, dream, wake, snooze, blanket, doze, slumber, snore, nap, peace, yawn, drowsy

THIEF: steal, robber, crook, burglar, money, cop, bad, rob, jail, gun, villain, crime, bank, bandit, criminal

Appendix B. DRM word list used in Chapter 4 and 5. Taken from Brainerd, Holliday, Reyna, Yang, and Toglia (2010).

Positive

HUG: kiss, tight, squeeze, embrace, affection, hold, cuddle, arms, bear, caress, touch, warmth

LOVE: like, happy, heart, care, admire, adore, close, friendship, happiness, kindness, life, girl

MUSIC: note, sound, piano, sing, art, band, melody, horn, concert, instrument, symphony, orchestra

NICE: sweet, good, kind, pleasant, friendly, smile, nasty, mean, polite, friend, people, bad

RIVER: water, stream, lake, boat, tide, swim, flow, run, creek, brook, fish winding

SOFT: light, pillow, plush, loud, cotton, fur, downy, fluffy, feather, furry, kitten, tender

BEACH: sand, sun, ocean, towel, ball, summer, surf, fun, wave, blanket, bum, coast

BIRD: nest, parrot, eagle, animal, wings, canary, cat, dog, flight, bluejay, tree, robin

SLEEP: bed, rest, awake, tired, dream, wake, snooze, doze, slumber, snore, nap, peace

PRETTY: beautiful, cute, adorable, attractive, flower, gorgeous, hot, model, pink, appealing, charming, lovely

Negative

HURT: pain, cut, ache, cry, injury, sad, ouch, fall, harm, hit, sore, stop

NEEDLE: thread, pin, eye, sewing, knitting, point, prick, thimble, thorn, injection, syringe, cloth

SICK: ill, well, hospital, vomit, doctor, flu, illness, fever, medicine, nausea, virus, pills

SPIDER: web, insect, bug, fright, arachnid, crawl, tarantula, poison, bite, creepy, feelers, small

COLD: snow, frost, winter, ice, wet, frigid, chilly, weather, freeze, air, shiver, arctic

DEAD: gone, funeral, bury, end, grave, asleep, still, alive, cemetery, deceased, die, duck

DIRT: mud, clean, soil, dirty, filth, ground, black, grime, unclean, brown, grease, mess

GLASS: window, cup, drink, break, clear, jar, crystal, fragile, house, bottle, plastic, sharp
FAT: skinny, big, ugly, thin, cow, slob, pig, diet, large, albert, lady, man
TRASH: garbage, waste, refuse, sewage, bag, junk, rubbish, sweep, scraps, litter, dump, landfill

Neutral

CHAIR: table, sit, legs, seat, couch, desk, recliner, sofa, cushion, stool, sitting, bench
FOOT: shoe, hand, toe, kick, sandals, walk, ankle, arm, boot, sock, knee, mouth
HIGH: low, clouds, up, tall, tower, jump, above, building, cliff, sky, over, elevate
FRUIT: apple, vegetable, orange, kiwi, citrus, bowl, pear, banana, basket, orchard, strawberry, grape
MOUNTAIN: hill, valley, climb, summit, top, peak, glacier, bike, climber, range, steep, ski

Appendix C. Metaphorical word pairs used in Chapter 6.

Conventional Word Pairs

Positive:

Warm heart: A kind, generous person.

Handsome profit: A large sum of money made

Bright student: A high achiever

Heavenly coffee: Delicious tasting coffee

Spotless record: Free from flaws

Neutral:

Purring motor: A low, vibrating sound from the engine

Dim recollection: A memory that is not clear in the mind

Rough estimate: An amount that you guess or estimate

Hard evidence: Facts that are definitely true and can be proven

Quiet life: Free from lots of work, not busy

Negative:

Clingy child: A child that is very attached and dependent

Thumping hangover: A severe headache caused by drinking too much

Cool reception: An unwelcoming greeting

Stinging betrayal: Pain felt when someone misleads or betrays

Transparent lie: A lie that is easily detected

Positive Congruent/Incongruent word pairs:

Set A

Glistening vocals: Good singing ability/Elaborate and exaggerated speech

Crackling excitement: Anticipation before an exciting event/Good singing ability

Audible confetti: Lots of different noises/A beautiful, colourful sunset

Glittery phrasing: Elaborate and exaggerated speech/Lots of different noises

Silky sunset: A beautiful, colourful sunset/Anticipation before an exciting event

Set B

Spicy stories: Exciting gossip about promiscuity/An overwhelming sense of happiness

Shiny bonuses: A generous reward/An image that makes you happy and joyful

Jingling satisfaction: Feeling very pleased and proud of yourself/A generous reward

Giggling picture: An image that makes you happy and joyful/Feeling very pleased and proud of yourself

Tingly joy: An overwhelming sense of happiness/Exciting gossip about promiscuity

Neutral Congruent/Incongruent word pairs:

Set A

Bleeping offices: A busy office environment/Distorted, unclear vision

Dusty dreams: A dream that is difficult to recollect/A busy office environment

Misty luck: Occasional good luck for an unknown reason/A dream that is difficult to recollect

Cold statistics: Presenting the statistics in a basic, unemotional way/Occasional good luck for an unknown reason

Muffled view: Distorted, unclear vision/Presenting the statistics in a basic, unemotional way

Set B

Tepid applause: An unenthusiastic, forced clapping/A poorly built house

Buzzing wave: Getting a sudden rush of energy/An unenthusiastic, forced clapping

Rhythmic climate: Regular seasonal changes in the weather/Getting a sudden rush of energy

Clicking attention: Sporadic interest with difficulties paying attention/Regular seasonal changes in the weather

Paper house: A poorly built house/Sporadic interest with difficulties paying attention

Negative Congruent/Incongruent word pairs:

Set A

Pale personality: A boring and dull person/Disapproval from any people

Slippery university: A failing university that is untrustworthy/A boring and dull person

Cruel building: Large, dark, uninviting architecture/A failing university that is untrustworthy

Elastic morals: Principles that change according to the situation/Large, dark, uninviting architecture

Echoing disapproval: Disapproval from many people/Principles that change according to the situation

Set B

Scratchy reaction: An irritated and hostile response/An uncomfortable and negative social situation

Thunderous collusion: A loud of forceful impact/An irritated and hostile response

Filthy morality: A lack of conscience/A loud and forceful impact

Snarling traffic: Busy traffic with lots of angry drivers/A lack of conscience

Chilly awkwardness: An uncomfortable and negative social situation/Busy traffic with lots of angry drivers