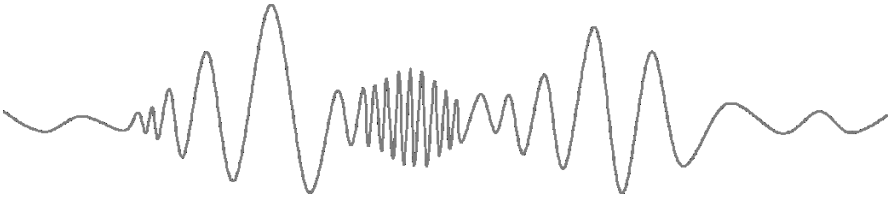
Sleep's role in the reprocessing and restructuring of memory

Anne Catharina Maria Koopman



A thesis submitted for the degree of Doctor of Philosophy

Cardiff University

School of Psychology

July 2020

Thesis Summary

Sleep disconnects us from our external environment and puts us in a vulnerable state, yet it is surprisingly universal. This thesis looks at the cognitive functions of sleep; specifically, the role of sleep in reprocessing and restructuring memory. It is now well-known that sleep actively consolidates memories, and even restructures them. This is likely achieved through the reactivation of memory representations. Previous research has shown that such reactivations can be triggered with a method called targeted memory reactivation (TMR).

In **Chapter 2**, I used TMR during rapid eye movement (REM) and slow-wave sleep (SWS) to investigate the effect of cueing in these stages on electrophysiology and subsequent task behaviour in a two-handed serial reaction time task. TMR during SWS led to detectable memory reactivation, and significant behavioural improvements in the non-dominant but not the dominant hand. TMR during REM did not affect behaviour, although electrophysiological results indicated that cues were processed during this stage. **Chapter 3** examined the effects of REM and SWS TMR on an associative memory task. We did not find any effect of SWS TMR. On the other hand, REM TMR improved remote associations between items which were not learned together but whose relationship could be inferred, indicating a role for REM sleep in memory restructuring. This was supported by a difference in event-related potentials in response to memory-related and control cues. However, two replications of the REM group showed that these results were not reliable. **Chapter 4**, finally, looked at the effects of wakefulness and sleep on two creative tasks. The more word-based task indeed benefitted from an interval containing sleep, but the more conceptual task showed improvements relating to wakefulness and time of day.

Together, these results increase our understanding of how different sleep stages, wakefulness, and memory reactivation all influence the restructuring of memory.

Declaration and Statements

DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is it being submitted concurrently for any other degree or award (outside of any formal collaboration agreement between the University and a partner organisation).

Signed: Anne Koopman____ Date: 09/07/2020

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD.

Signed: Anne Koopman____ Date: 09/07/2020

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated, and the views expressed are my own. Other sources are acknowledged by explicit references. The thesis has not been edited by a third party beyond what is permitted by Cardiff University's Use of Third Party Editors by Research Degree Students Procedure.

Signed: Anne Koopman____ Date: 09/07/2020

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available in the University's Open Access repository (or, where approved, to be available in the University's library and for inter-library loan), and for the title and summary to be made available to outside organisations, subject to the expiry of a University-approved bar on access if applicable.

Signed: Anne Koopman____ Date: 09/07/2020

WORD COUNT: 59,311

(Excluding summary, acknowledgements, declarations, contents pages, appendices, tables, diagrams and figures, references, bibliography, footnotes and endnotes)

Table of Contents

Thesis Summary v
Declaration and Statements vii
Acknowledgements xiii
Contributors xv
CHAPTER 1: General introduction 17
1.1 Preface 18
1.2 Sleep physiology 19
1.3 Sleep and Memory 21
1.3.1 Memory processes and systems 22
1.3.2 Models of sleep & memory 23
1.4 Memory reactivation and Targeted Memory Reactivation (TMR) 26
1.4.1 Spontaneous reactivation 26
1.4.2 Targeted Memory Reactivation (TMR) 29
1.4.3 Detection of human (sleep) reactivations 31
1.5 The restructuring of memories during sleep 34
1.5.1 Regularity abstraction and generalisation 34
1.5.2 Associative inference 36
1.5.3 Creativity 37
1.6 Summary 41
1.7 Research objectives 42
CHAPTER 2: Targeted memory reactivation of a serial reaction time task in SWS, but not REM, preferentially benefits the non-dominant hand 45
2.1 Abstract 46
2.2 Introduction 46
2.3 Materials and Methods 49
2.3.1 Participants 49
2.3.2 Experimental tasks and design 49
2.3.3 PSG data acquisition and analysis 53
2.3.4 Electrophysiological analysis 54
2.3.5 Spindle analysis 55
2.3.6 Behavioural analysis 56
2.3.7 Classification 58
2.4 Results 60
2.4.1 Sleep parameters 60

2.4.2 Behaviour: Both Hands 61
2.4.3 Behaviour: Left and Right Hand 62
2.4.4 Electrophysiology: Spindle analysis 63
2.4.5 Electrophysiology: Event-related potentials 65
2.4.6 Classification 66
2.4.7 Correlating classification performance with behaviour 69
2.5 Discussion 70
2.5.1 TMR in SWS but not REM benefits SRTT consolidation 71
2.5.2 TMR preferentially benefits the non-dominant hand 72
2.5.3 Linear classifier with time domain features detects reactivation in SWS 73
2.5.4 Conclusion 74
2.6 Supplements 75
2.6.1 Explicit Recall 75
2.6.2 Electrophysiology: Event-related potentials per night 75
2.6.3 Time-frequency analysis 75
CHAPTER 3: The effect of targeted memory reactivation in REM and SWS on remote associations 81
3.1 Abstract 82
3.2 Introduction 82
3.3 Experiment 1: Materials and Methods 85
3.3.1 Participants 85
3.3.2 Purpose 86
3.3.3 Stimuli 86
3.3.4 Experimental protocol 87
3.3.5 Targeted memory reactivation 90
3.3.6 Polysomnography (PSG) data acquisition and analysis 91
3.3.7 Additional tasks 92
3.3.8 Behavioural data analysis 92
3.3.9 EEG data analysis 93
3.4 Experiment 1: Results 94
3.4.1 Sleep data 94
3.4.2 Learned associations (face – scene) 95
3.4.3 Remote associations (face – face) 97
3.4.4 Event-related potential analysis 99
3.4.5 Time-frequency analysis 100
3.5 Experiment 1: Discussion 102

3.6 Experiment 2: Materials and Methods 103
3.6.1 Participants 103
3.6.2 Experimental protocol 104
3.6.3 Targeted memory reactivation 104
3.6.4 Polysomnography (PSG) data acquisition and analysis 105
3.6.5 Data analysis 105
3.7 Experiment 2: Results 105
3.7.1 Sleep data 105
3.7.2 Learned associations (face-scene) 106
3.7.3 Remote associations (face-face) 107
3.7.4 Event-related potential analysis 108
3.7.5 Time-frequency results 109
3.8 Experiment 2: Discussion 110
3.9 Experiment 3: Materials and Methods 111
3.9.1 Participants 111
3.9.2 Experimental protocol 112
3.9.3 Targeted memory reactivation 112
3.9.4 Polysomnography (PSG) data acquisition and analysis 112
3.9.5 Data analysis 113
3.10 Experiment 3: Results 113
3.10.1 Sleep data 113
3.10.2 Learned associations (face-scene) 114
3.10.3 Remote associations (face-face) 115
3.10.4 Extra behavioural analyses with probe switched 116
3.10.5 Event-related potential analysis 117
3.10.6 Time-frequency analysis 118
3.10.7 Combined analyses 119
3.11 Discussion 125
3.11.1 TMR during SWS does not improve learned associations 126
3.11.2 TMR during REM sleep does not lead to consistent effects 127
3.11.3 Possible explanations for the disparity between the REM groups 128
3.11.4 The wider context of sleep effects on remote associations 130
3.11.5 Conclusion 131
CHAPTER 4: The effect of sleep and wakefulness on creativity 133
4.1 Abstract 134
4.2 Introduction 134

4.3 Materials and Methods 137
4.3.1 Participants 137
4.3.2 Experimental Protocol 138
4.3.3 Verb Generation Task (VGT) 139
4.3.4 Alternative Uses Task (AUT) 141
4.3.5 Psychomotor Vigilance Test (PVT) 142
4.3.6 Data Analysis 142
4.4 Results 146
4.4.1 Sleep 146
4.4.2 Questionnaires 146
4.4.3 Verb Generation Task (VGT) 147
4.4.4 Alternative Uses Task (AUT) 151
4.4.5 Psychomotor Vigilance Test (PVT) 154
4.5 Discussion 155
4.5.1 Sleep leads to higher semantic distance in the VGT than wakefulness 156
4.5.2 An over-day interval improves performance on the AUT 158
4.5.3 Possible explanations for the disparity between the AUT and VGT 160
4.5.4 Relationships with alertness and sleep parameters 161
4.5.5 Conclusion 163
CHAPTER 5: General discussion 165
5.1 Overview 166
5.2 Experiments in this thesis 167
5.2.1 Summary of findings 167
5.2.2 Limitations of the experiments 168
5.3 Memory reprocessing during sleep 172
5.3.1 Reprocessing during SWS 172
5.3.2 Reprocessing during REM 174
5.3.3 The selectivity of reprocessing 177
5.4 Memory restructuring during sleep 179
5.4.1 Does sleep really promote memory restructuring? 179
5.4.2 The roles of wakefulness, REM, and NREM sleep 181
5.5 The link between reprocessing and restructuring 183
5.6 Future directions 187
5.7 Conclusion 188
References 191
Appendices 223

Acknowledgements

This thesis would not have been possible without the help, support, guidance, friendship, and love of a great number of people, and I would like to thank them deeply for everything that they have done. First, I want to thank my supervisor, Penny Lewis, for her enthusiasm, expertise, and feedback throughout the years that we worked together. Thanks also go to Matthias Gruber and Rob Honey, who read parts of this thesis and helped improve the work and writing with their suggestions.

Very important help and friendship was provided by the NaPS lab, past and present. In particular, I want to thank (in alphabetical order) Alexia, Alun, Duarte, Holly, Jen, Jules, Karen, Lorena, Mahmoud, Martyna, Matthias Treder, Miguel, Natalie, Paulina, Shi Wei, and Sofia. In truth, I benefitted from the friendship and support of so many of my fellow researchers in CUBRIC and the School of Psychology. I am very grateful to have had the opportunity to learn from you. I should also thank the School of Psychology at Cardiff University itself, for funding this PhD.

I was lucky enough to meet a wonderful group of PhD researchers during a workshop in Budapest in my first year. I owe my thanks to this whole group of lovely people, but especially to Maggie Webb for her help with creativity-related questions. David Schöntal and Andrew Martin, my housemates during my first year, showed me that it is important to do things besides work. There were also various people on Twitter (you know who you are) who encouraged me when I was feeling low, for which I was and am very grateful. Special thanks go to all of the participants who made it possible for me to do this work: thank you for allowing me a look inside your sleeping brain! I should give credit where credit is due, which means I have to thank the makers of all the TV shows I watched during my many nights of collecting data, without which I would definitely have fallen asleep. Thanks also go to all the snacks and cups of tea that kept me sane(ish) during pandemic thesis writing.

Finally, I want to thank my family, for always being supportive and even reading (some of) my work sometimes! Major thanks go to Paulien, who always knows how to cheer me up and who sent me a box of liquorice at somehow exactly the right moment. Above all, of course, I thank Timo, for being there through all of it. I am sure neither of us expected me to finish my thesis while you were working half a metre away from the same kitchen table, but I could not have wished for a better partner in this journey.

Contributors

Sleep research is a team effort. Not only is the presence of another researcher during an overnight study an institutional requirement, but the help of various people, especially with data collection, allowed my health to be somewhat unaffected by the various months of night shifts. I owe my thanks to all those who contributed to this research, whose names are mentioned below specifically for each chapter. Here, I focused on contributors outside of the supervisory team.

In **Chapter 2**, Monika Śledziowska helped me convert existing task scripts to fit my design. Participant recruitment and data collection (including overnight cueing) was shared between Monika Śledziowska, Thomas Hunter, and myself. Alun Metcalf and I sleep scored the data. Dr Suliman Belal wrote scripts to convert the EEG data into a usable format. Martyna Rakowska wrote scripts to extract the behavioural data from the results files. EEG artifact rejection and spindle detection scripts were based on scripts written by Miguel Navarrete. All analyses were done by me, with the exception of the classifier analysis, which was developed and carried out by Mahmoud Abdellahi. I wrote up the entire chapter, but the classifier sections were written jointly with Mahmoud Abdellahi. This is also explicitly acknowledged at the start of Chapter 2.

The control group in Experiment 1 of **Chapter 3** was largely collected by Nora Hennies and Marleen Kempkes at Manchester University. Nora was also involved in the design of the study, together with myself, and she wrote the task scripts. All recruitment for Experiments 1 (SWS and REM groups) and 3 was done by me, and recruitment for Experiment 2 was done by Natalie Gunasekara. Karen Konkoly and I collected the data for Experiment 1 (SWS and REM groups, and four control group participants). Duarte Pereira, Shi Wei Teo, Natalie Gunasekara, and I collected the data for Experiment 2. Paulina Bagrowska, Ibad Kashif, and Niall McGinley helped with data collection of Experiment 3, though I did all the overnight cueing for this experiment. Duarte Pereira, Shi Wei Teo, Elena Schmidt, and I sleep scored the data of Experiments 1 and 2. The EEG artifact rejection script was based on a script written by Miguel Navarrete. I analysed the data and wrote the chapter.

Design of the protocol of **Chapter 4** was done together with Martyna Rakowska. Recruitment and data collection was shared between Martyna Rakowska, Ralph Andrews, and myself. Marit Petzka wrote the PVT script. I wrote the VGT and AUT scripts. Dr Maggie Webb provided the AUT scoring form, and guidance on scoring the AUT. I analysed the data and wrote the chapter.

CHAPTER 1

General introduction

1.1 Preface

Every night, when the sun goes down and night falls, something rather strange happens: people start to yawn, retreat to their bedrooms, and collectively lose consciousness. We become largely unresponsive to our external environment, and remain in this vulnerable state until morning. Then, we wake up, and we go about our business as if this is all the most normal thing in the world. Stranger still, humans are not the only ones that display this type of behaviour. Animals, and even plants and certain bacteria, also show circadian rhythms with alternating periods of activity and rest (Golden, Ishiura, Johnson, & Kondo, 1997; McClung, 2006). Many species, indeed, become as unresponsive during sleep as humans (Siegel, 2008). The question we are still trying to solve is, why?

Sleep can be defined as "a rapidly reversible state of immobility and greatly reduced sensory responsiveness" (Siegel, 2008, p. 208). Moreover, sleep is homeostatically regulated – meaning that a reduction in sleep is subsequently followed by an increased need for sleep ('sleep rebound'). Although sleep appears to be fundamental for many species, many questions remain about its function. Historically, sleep has been seen as a passive state, a state of mere inactivity, but starting with the first electroencephalographic (EEG) recording of sleep in 1924 this has slowly changed (Haba-Rubio & Krieger, 2012). For the first time, sleep could be measured and studied objectively, and recordings of full nights of sleep thereafter revealed that the brain appears to be quite busy while asleep. Sleep is characterised by different stages and oscillatory patterns, and it is quite possible that each of these patterns serve different or complementary functions.

In this thesis, the focus will be on the cognitive functions of sleep, specifically the role of sleep in the reprocessing and restructuring of memory. As memories consolidate, they are thought to be reorganised into neocortical networks (McClelland, McNaughton, & O'Reilly, 1995). In other words, memories are not only strengthened, but also integrated, transformed, and restructured. There is now compelling evidence that sleep plays an active role in this process (Diekelmann & Born, 2010; Rasch & Born, 2013).

One of the key mechanisms through which sleep has been proposed to fulfil these functions is through the reactivation or replay of memory representations. In rodents, specific patterns of neuronal firing that occurred during learning were found to reoccur during subsequent sleep (Skaggs & McNaughton, 1996; Wilson & McNaughton, 1994), almost as if the brain was rehearsing

what it had learned. Indeed, in the years since these influential papers, ample evidence has been brought forward establishing a role of post-learning reactivation during sleep in human memory consolidation (e.g. Bergmann, Mölle, Diedrichs, Born, & Siebner, 2012; Cairney, Guttesen, El Marj, & Staresina, 2018; Maquet et al., 2000; Peigneux et al., 2004; Schönauer et al., 2017). Nowadays, many studies make use of a technique called targeted memory reactivation (TMR). TMR pairs sensory cues such as sounds or smells with a learning task given to participants before they sleep. Then, as the participant sleeps, the sensory cues are re-presented, which can bias memory reactivation and subsequent memory consolidation and lead to behavioural changes (Rasch & Born, 2013; Rasch, Büchel, Gais, & Born, 2007).

The aim of this thesis is to add to the current understanding of the role of sleep in the reprocessing and restructuring of memory, looking at three different kinds of tasks and mainly using TMR. In this general introduction, the relevant background will be presented, starting with the current understanding of sleep physiology. Then, I will take a closer look at the link between sleep and memory, discussing the ways in which memory has been conceptualised and models that have included a key role for sleep in memory. Furthermore, I will dive deeper into the concepts of memory reactivation and TMR. Special attention will be paid to the restructuring of memories during sleep, looking at different ways in which this may present itself. This introduction concludes with a summary of the most important points and the presentation of my research objectives.

1.2 Sleep physiology

As mentioned above, sleep outwardly consists of several characteristics, such as a reduction in responsiveness to external stimuli, comparative inactivity, and a loss of consciousness. Methods such as polysomnography (PSG), which combines electroencephalography (EEG), electrooculography (EOG) and electromyography (EMG), enable us to take a closer look. When we do this, it becomes clear that sleep is not uniform, but instead consists of several stages (see Figure 1.1A).

First, two major sleep stages can be distinguished: rapid eye movement sleep (REM) sleep, and non-REM (NREM) sleep, the latter of which can further be divided into four stages, stages 1-4. At the start of a night of sleep, people often spend a short amount of time in Stage 1 (S1). This is a transitional stage, which typically only constitutes up to 10% of total sleep time (TST) (Moser et al.,

2009). Stage 1 sleep is characterised by a slowing of eye movements and a reduction in the amount of alpha (8-12 Hz) waves compared to wake (Silber et al., 2007).

This stage is usually followed by Stage 2 (S2), which makes up the majority of the night of sleep – about 45-55%. In this stage, one finds K-complexes and spindles. K-complexes consist of a negative sharp wave followed by a large positive component. They can be spontaneous, but also occur after a sudden noise in the environment, which has been interpreted as reflecting a role in the maintenance of sleep (Cash et al., 2009). Spindles, in turn, are short bursts of relatively high-frequency activity. There is some disagreement about the exact frequency of spindles, but they are generally considered to be in the 10-15 Hz range, and often divided into slow (10-12 Hz) and fast (13-15 Hz) subgroups (Barakat et al., 2011; Fernandez & Lüthi, 2020; Mölle, Bergmann, Marshall, & Born, 2011). Sleep spindles have been linked to learning (Fernandez & Lüthi, 2020; Peyrache & Seibt, 2020; Ulrich, 2016).

Sleep deepens further into slow-wave sleep (SWS), a stage which was formerly divided into Stages 3 and 4. This stage makes up around 15-20% of the TST, predominantly in the first half of the night. It is characterised by slow oscillations (<1 Hz) and delta waves (1-4 Hz), giving it its typical wave-like appearance (see Figure 1.1A, lowest EEG trace). Spindles still occur during SWS, albeit less frequently than during Stage 2. SWS, too, has been linked to memory consolidation (Diekelmann & Born, 2010; Navarrete, Valderrama, & Lewis, 2020; Rasch & Born, 2013).

The last of the sleep stages is REM sleep, which takes up approximately 20-25% of TST. Due to its rapid eye movements and low-amplitude, high frequency nature, the EEG trace in this stage resembles that of wakefulness. REM sleep has therefore also been called ‘paradoxical sleep’. Other important components of REM sleep include sawtooth waves (2-6 Hz) and muscle atonia as seen in the EMG. The combination of all these characteristics make it possible to correctly distinguish this stage. Automatic sleep scoring algorithms often have particular difficulty recognising REM sleep, which means that sleep scoring is still mostly done manually to obtain the most reliable results. However, recent advancements have led to high-accuracy automated alternatives (e.g. Patanaik, Ong, Gooley, Ancoli-Israel, & Chee, 2018).

In humans, sleeping is usually done at night in periods of approximately 8 hours. Within these 8 hours, people cycle through the different sleep stages in roughly 90 minutes. The first part of the night is dominated by NREM sleep, whereas later parts of the night often contain more REM sleep and little to no SWS (illustrated by the hypnogram in Figure 1.1A). In contrast, mice and rats, often-

used animal models in sleep research, sleep predominantly during the day. Moreover, they cycle through sleep stages much more rapidly than humans (Figure 1.1B), and animal researchers usually do not distinguish between the different stages within NREM sleep. Despite these differences, key sleep features like spindles and slow waves are relatively similar to those found in humans, making rodents an appropriate model for studying the relationship between sleep and memory (Datta & Hobson, 2000; Datta & MacLean, 2007; Doran, Wessel, Kilduff, Turek, & Renger, 2008; Tobler, Franken, Trachsel, & Borbély, 1992; Veasey et al., 2000; Yasenkov & Deboer, 2010).

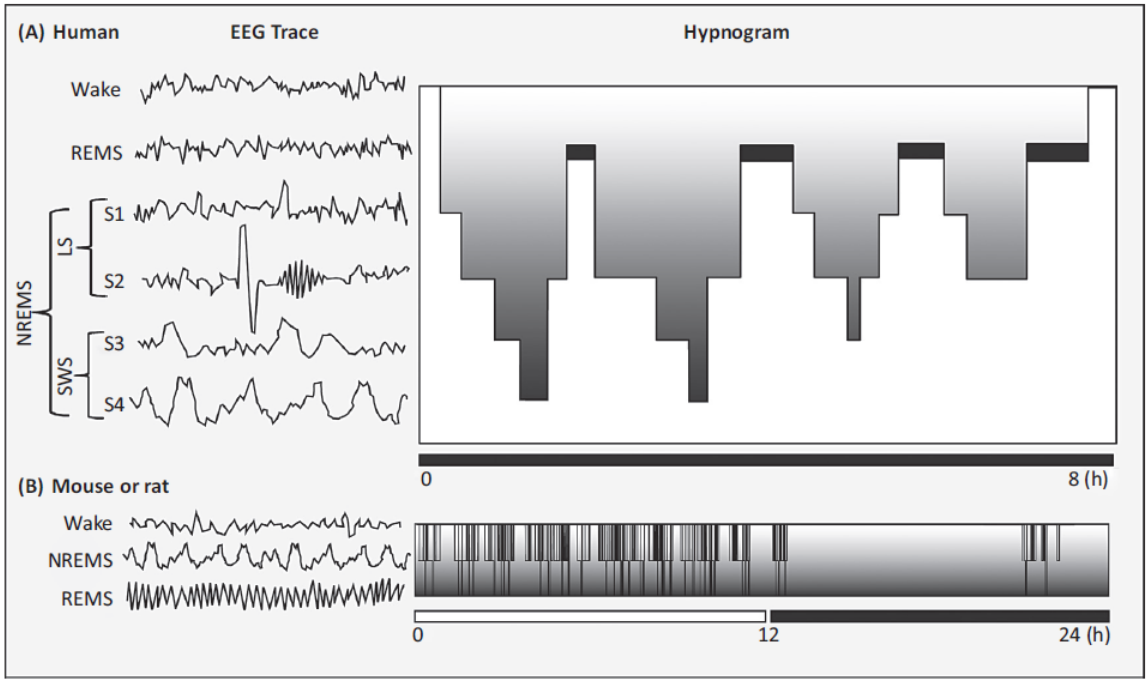


Figure 1.1. *EEG Trace and Hypnogram of sleep in (A) humans and (B) mice and rats. Figure reproduced with permission from* Genzel, Kroes, Dresler, & Battaglia (2014).

1.3 Sleep and Memory

To adapt our behaviour, make predictions about the future, and learn from the past, it is crucial that we remember it. Memory, then, is one of our most important abilities. Given the vast amount of work that has been conducted on this topic, it would be impossible to cover it comprehensively. Instead, this section is meant to be a short overview of important concepts, particularly as they relate to sleep.

1.3.1 Memory processes and systems

The concept of memory can be divided into three stages: encoding, consolidation, and retrieval. Encoding is the process by which a new memory trace is formed, following the perception of a stimulus. Initially, these new memories are labile, but through consolidation they become more stable and are reorganised into the existing long-term memory network. The last stage, retrieval, is the remembering of a memory. Recent work has indicated that encoding may be possible during sleep under certain circumstances (Andrillon, Pressnitzer, Léger, & Kouider, 2017; Arzi et al., 2012; De Lavilléon, Lacroix, Rondi-Reig, & Benchenane, 2015; Züst, Ruch, Wiest, & Henke, 2019). However, consolidation is where the role of sleep becomes most apparent. There is a wealth of evidence supporting the idea that sleep benefits memory consolidation, much of which will be discussed throughout this introduction (Diekelmann & Born, 2010; Rasch & Born, 2013; Stickgold, 2005).

Memories are commonly divided into multiple different types, the major distinction being between declarative and nondeclarative memory (Cohen & Squire, 1980; Squire, 2004). Declarative memory is generally considered to be accessible by conscious recall. Moreover, the encoding of this type of memory is thought to depend on brain structures in the medial temporal lobe, particularly the hippocampus (Squire & Zola, 1996). On the other hand, the nondeclarative memory branch is composed of things like motor skills, conditioning, and perceptual skills. These are thought to be hippocampus-independent during encoding and can be acquired and retrieved without awareness.

Declarative memories are further divided into episodic and semantic memories, the former encompassing memories for events which contain a spatial and temporal context (Tulving, 1983). Semantic memories, in contrast, are facts that we remember without such context. This latter category also includes the representation of language and language comprehension (Binder & Desai, 2011; Kutas & Federmeier, 2000).

These distinctions between declarative-nondeclarative and episodic-semantic provide a useful way of talking about memory, especially in an experimental context. However, it is important to note that learning does not always follow these divisions. For instance, procedural tasks and language learning often include explicit and implicit components, the involvement of which changes over the course of acquisition (Doyon & Benali, 2005; Peigneux, Laureys, Delbeuck, & Maquet, 2001). The hippocampus, not thought to be implicated in nondeclarative memory, has also been shown to be involved in the training of certain motor tasks (Albouy et al., 2008). Additionally, the semantic-

episodic distinction, though it often proves useful, has been criticised from the start (Anderson & Ross, 1980; McKoon, Ratcliff, & Dell, 1986) and recent research has highlighted ways in which they are intertwined (Irish & Piguet, 2013; Renoult, Irish, Moscovitch, & Rugg, 2019).

Despite the categorisation of this wide variety of different memory types, consolidation during sleep appears to be involved in all of them (Stickgold, 2005). This involvement may be explained through a discussion of the currently dominant theory of human memory: the two-stage memory system (Marr, Willshaw, & McNaughton, 1971; McClelland et al., 1995). At the core of this system is the idea that memories, when they are first encoded, are saved in a fast learning store. For example, in the case of declarative memories, this would be the hippocampus. This fast learning store allows memories to be quickly encoded, but within this store they are labile. Then, over time, some new memories are moved to a long-term store (e.g. the cortex). This leads to their stabilisation, integration with existing memories, and even their reorganisation (Dudai, Karni, & Born, 2015). This move to the long-term store is believed to occur through the repeated reactivation of new memories during rest periods, including sleep (Rasch & Born, 2013).

1.3.2 Models of sleep & memory

The idea that sleep plays an important role in memory consolidation thus fits very well within the theory of the two-stage memory system. Long before the formulation of this theory, however, it had already been discovered that a period of sleep reduces forgetting compared to wakefulness (Heine, 1914; Van Ormer, 1933). Early research focused on the role of sleep as passive protector, which reduced forgetting because the encoding of new information, and subsequently interference, was minimised (Rasch & Born, 2013; Wixted, 2004). However, this view that sleep plays a passive role in memory consolidation was not compatible with findings that a different composition of sleep in terms of sleep stages was associated with different results in terms of memory retention (Barrett & Ekstrand, 1972; Fowler, Sullivan, & Ekstrand, 1973; Plihal & Born, 1997; Yaroush, Sullivan, & Ekstrand, 1971).

A popular way of achieving sleep with a different sleep stage composition is the ‘night half paradigm’, wherein participants perform a memory task and sleep either in the first or in the second half of the night (Fowler et al., 1973; Yaroush et al., 1971). As mentioned in section 1.2, the first half of the night is predominantly made up of NREM sleep, and SWS in particular. The second half, on the other hand, is dominated by REM sleep and contains little to no SWS. Findings from these studies led to the formulation of the Dual Process Hypothesis, which posits that different types of

memories benefit from different sleep stages (Plihal & Born, 1997; Wagner, Gais, & Born, 2001). Specifically, it was suggested that procedural and emotional memories are enhanced by REM sleep, whereas declarative memories improved with SWS. However, subsequent experiments have not consistently demonstrated this distinction (e.g. Aeschbach, Cutler, & Ronda, 2008; Fogel, Smith, & Cote, 2007; Gais, Plihal, Wagner, & Born, 2000; Huber, Felice Ghilardi, Massimini, & Tononi, 2004; Rauchs et al., 2004). In addition, this model largely overlooks the influence of Stage 2 sleep on memory, which has been implicated in the learning of both procedural (Fogel & Smith, 2006; Laventure et al., 2016) and declarative tasks (Clemens, Fabó, & Halász, 2005, 2006; Ruch et al., 2012).

In contrast, the Sequential Hypothesis focuses on the interaction between the different sleep stages rather than distinguishing the types of memory that might benefit from a particular stage (Ambrosini & Giuditta, 2001; Giuditta, 2014; Giuditta et al., 1995). In particular, it hypothesises that the cyclical progression of sleep stages, including the alternation between SWS (or NREM) and REM sleep, is the key to memory consolidation. It considers SWS to be the stage that maintains certain (useful) memories and removes or downscales other, irrelevant or interfering, memories (Giuditta, 2014). Subsequently, REM sleep is responsible for strengthening the useful memories and integrating them with past memories. This hypothesis has been supported by several experiments with humans (Ficca & Salzarulo, 2004; Gais et al., 2000; Mazzoni et al., 1999; Stickgold, Whidbee, Schirmer, Patel, & Hobson, 2000). For instance, one study established that naps containing both SWS and REM sleep improved performance on a texture discrimination task, whereas naps containing only SWS did not (Mednick, Nakayama, & Stickgold, 2003). Another experiment showed that fragmented sleep during which sleep cycles were disrupted decreased recall of verbal material, but the same amount of fragmentation did not impair memory when the sleep cycles were maintained (Ficca, Lombardo, Rossi, & Salzarulo, 2000). Results of several experiments thus appear to be in line with this hypothesis. Nevertheless, direct tests of the ways in which SWS and REM interact have been largely absent, which means that there is not much evidence either for or against this model (Sara, 2017; Scullin & Gao, 2018).

Currently, the model that seems to receive the largest amount of support is the Active System Consolidation Hypothesis, which integrates aspects of the previous two models and focuses on the active role of sleep in memory consolidation (Diekelmann & Born, 2010). In this model, reactivations of new memory representations take on a central function. During encoding, new memories are saved in a temporary store – in the case of declarative memories, the hippocampus.

Afterwards, during SWS, these memory traces are reactivated repeatedly, and through this reactivation they are moved to their long-term store (e.g. the cortex). In the hippocampus, these reactivations are said to take place during sharp wave-ripples (Buzsáki, 2015). The temporal coupling of sharp-wave ripples, spindles, and slow oscillations during SWS is hypothesised to drive this active process of consolidation (Rasch & Born, 2013). The role of REM sleep in this process is thought to be to stabilise the transported memories, through processes of synaptic consolidation. Indeed, both SWS and spindles have repeatedly been associated with increased learning (Cairney et al., 2018; Gais, Mölle, Helms, & Born, 2002; Schabus et al., 2004). For instance, a pharmacologically induced increase in the amount of sleep spindles during a nap following a word-pair task led to greater memory improvement compared to a placebo (Zhang, Yetton, Whitehurst, Naji, & Mednick, 2020). Another recent study has further found evidence that sharp wave-ripples during NREM sleep are related to memory replay and consolidation in humans (Zhang, Fell, & Axmacher, 2018). Furthermore, increasing temporal coupling between the hippocampus and cortex led to benefits for memory consolidation in mice (Maingret, Girardeau, Todorova, Goutierre, & Zugaro, 2016).

As mentioned before, memories are thought to move from a short- to a long-term store through the repeated reactivation of these memories during rest periods (Rasch & Born, 2013). The models mentioned above provide hypotheses about how this could specifically be accomplished, and what role different sleep stages play in this process. However, not everyone agrees that memory reactivation is the driving force behind memory consolidation. A somewhat competing model is the Synaptic Homeostasis Hypothesis or SHY (Tononi & Cirelli, 2003). In this theory, the emphasis is put on the homeostatic function of sleep in regulating synaptic potentiation. It posits that during wake, synapses in the cortex become potentiated, but this cannot go on forever. The function of slow-wave activity is to downscale these synapses in order to reach synaptic homeostasis, ready for the next waking period. Through this downscaling, strong memories or those that are tagged as important are weakened less than already weak or irrelevant memories, achieving a higher signal-to-noise ratio which explains the beneficial effect of sleep on memory (Tononi & Cirelli, 2014).

SHY is often presented in the literature as an alternative to models that favour reactivation, and there is a competition between them. For instance, in the SHY view, spontaneous memory reactivation is unlikely to lead to any meaningful impact on consolidation without a large-scale synaptic downscaling which improves the signal-to-noise ratio for those reactivated memories (Tononi & Cirelli, 2014). On the other hand, SHY is often criticised for lacking an explanation for the

fact that downscaling also takes place during wake, and potentiation during sleep (Rasch & Born, 2013). Nevertheless, both sides to this debate acknowledge that there is experimental evidence that supports each theory. In fact, newer incarnations of SHY hypothesise that memory reactivation could work in conjunction with downscaling (Tononi & Cirelli, 2014, 2016). In this view, synapses that are reactivated strongly or often during sleep are protected from downscaling. Thus, SHY mainly offers a different interpretation of what the most important factor in memory consolidation during sleep is. Notably, there are also other models which combine elements of synaptic homeostasis and reactivation, such as the (B)iOtA model (Lewis, Knoblich, & Poe, 2018). This model is specifically concerned with memory restructuring and will be explained further in section 1.5.3.

1.4 Memory reactivation and Targeted Memory Reactivation (TMR)

The previous section outlines some of the ample evidence that sleep affects memory consolidation. Notably, although several models exist that aim to clarify exactly how sleep influences memory, the importance of memory reactivation is a part of most, if not all, of these models.

1.4.1 Spontaneous reactivation

The first and some of the most compelling evidence for memory reactivation is found in the rodent literature on place cells. Place cells are neurons in the hippocampus that account for an animal's position in its environment (Burgess, Donnett, & O'Keefe, 1998; O'Keefe, Nadel, & Willner, 1979). As the animal moves through space, different place cells fire depending on the spatial field in which the animal is located. Place field-related spiking activity that occurred during wakefulness has been shown to reoccur during subsequent sleep (Pavlides & Winson, 1989).

Even more convincingly is a series of studies in rats that recorded a large number of place cells during a task and during sleep (Skaggs & McNaughton, 1996; Wilson & McNaughton, 1994). As the rat ran along a track, certain cells fired together. During SWS which followed the task, these same cells had an increased tendency to fire together – something which was not seen during sleep before the task (Wilson & McNaughton, 1994). Importantly, these cells not only fired together, but showed a temporal order which was preserved during sleep (Skaggs & McNaughton, 1996). A visual representation of these studies can be found in Figure 1.2.

Reactivation during SWS has since been frequently demonstrated, mainly during sharp wave-ripples (Girardeau, Cei, & Zugaro, 2014; Kudrimoti, Barnes, & McNaughton, 1999; Nakashiba, Buhl,

McHugh, & Tonegawa, 2009; O'Neill, Senior, Allen, Huxter, & Csicsvari, 2008). It was found that temporal order is maintained, although the speed of reactivations seems to be much faster than that of activity during encoding (Hirase, Leinekugel, Czurkó, Csicsvari, & Buzsáki, 2001; Ji & Wilson, 2007; Lee & Wilson, 2002; Nádasdy, Hirase, Czurkó, Csicsvari, & Buzsáki, 1999; Skaggs, McNaughton, Wilson, & Barnes, 1996). Moreover, reactivation mainly appears to occur for a short period following learning, after which it is reduced, which potentially indicates that there is a limit to the amount of reactivations that are useful or necessary for consolidation (Battaglia, Sutherland, Cowen, Mc Naughton, & Harris, 2005; Kudrimoti et al., 1999; Qin, Mcnaughton, Skaggs, & Barnes, 1997; Shen, Kudrimoti, McNaughton, & Barnes, 1998; Skaggs et al., 1996).

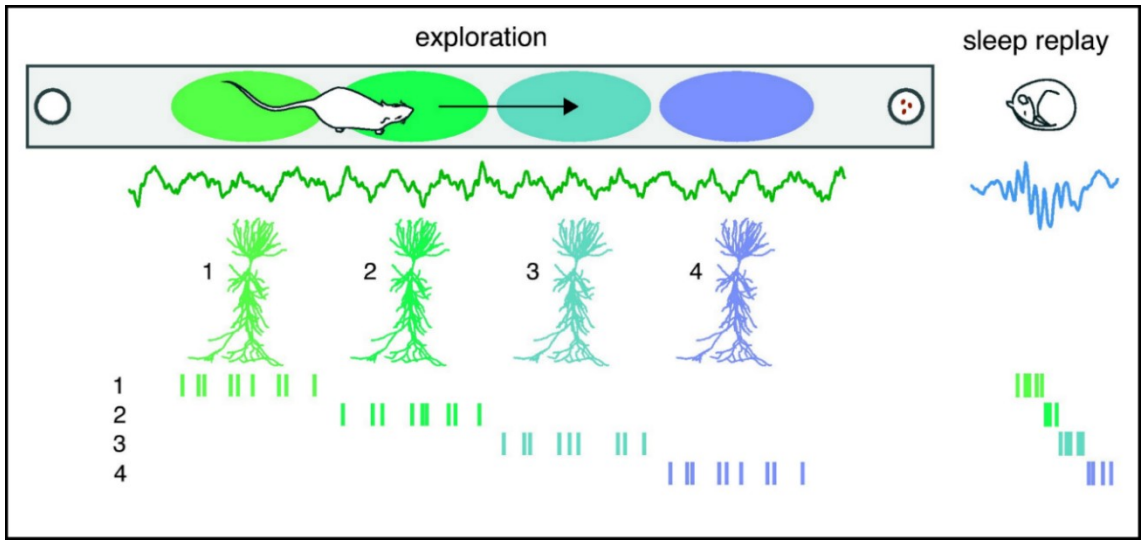


Figure 1.2. *Neuronal replay in sleep. As a rat runs along a track, hippocampal place cells are activated in succession as the rat enters their place fields (coloured ellipses). This results in a neuronal sequence (vertical ticks). Afterwards, when the rat is asleep or in quiet rest, these place cells reactivate in the same order. Figure adapted with permission from* Girardeau & Zugaro (2011).

Providing support for the Active System Consolidation Hypothesis, reactivation during sleep has also been shown in structures beyond the hippocampus. For example, reactivation has been observed in several cortical areas, like the parietal (Qin et al., 1997), prefrontal (Euston, Tatsuno, & McNaughton, 2007; Johnson, Euston, Tatsuno, & McNaughton, 2010; Peyrache, Khamassi, Benchenane, Wiener, & Battaglia, 2009), and visual cortices (Ji & Wilson, 2007). In fact, this latter study found that reactivation followed a temporal pattern, occurring slightly earlier in the hippocampus than the visual cortex. This was corroborated by studies looking at a subcortical area called the ventral striatum, where reward-related information appeared to be reactivated (Pennartz et al., 2004), which was again preceded by hippocampal reactivations (Lansink et al., 2008; Lansink, Goltstein, Lankelma, McNaughton, & Pennartz, 2009). Several of these studies show

NREM sleep reactivation of non-spatial information, indicating that many features (i.e. not just spatial) of an animal's behaviour during wakefulness undergo this process.

Rodent evidence of reactivation during REM sleep is more limited, although this may partly be due to the fact that few studies have looked at this stage. Nevertheless, the very first study looking at the reoccurrence of place field-related spiking activity during sleep found this during REM as well as SWS (Pavlides & Winson, 1989). Another study showed that the firing phase of hippocampal neurons during REM theta oscillations depended on the familiarity of the environment explored during wake (Poe, Nitz, McNaughton, & Barnes, 2000). Louie and Wilson (2001) demonstrated that temporal sequences of wake activity in multiple neurons in the hippocampus were reinstated during REM sleep, and interestingly the timescale was similar to that of wakefulness. Recently, REM replay was also found in the primary visual cortex (Howe, Wilson, Ji, & Jones, 2019).

It is worth noting that reactivation not only occurs during sleep, but also in wakefulness, particularly during rest periods (Carr, Jadhav, & Frank, 2011; Tambini & Davachi, 2019). This awake reactivation has also been linked to memory consolidation, but it is unclear whether it serves similar or complementary functions to reactivation during sleep.

Various studies have shown a beneficial effect of spontaneous memory reactivation on memory consolidation. In rodents this is generally difficult to demonstrate because the animals are so thoroughly trained on the tasks that they no longer improve. Nonetheless, one study did find that reactivation of goal-related task information during rest after learning was associated with subsequent memory performance (Dupret, O'Neill, Pleydell-Bouverie, & Csicsvari, 2010). Furthermore, the disruption of sharp wave-ripples during post-learning rest periods has been shown to have a detrimental effect on spatial memory (Girardeau, Benchenane, Wiener, Buzsáki, & Zugaro, 2009; Girardeau & Zugaro, 2011). Even more directly, in an experiment that used training of two different environments and online reactivation decoding, selective disruption of the reactivation of one of the environments led to performance impairments in that environment (Gridchyn, Schoenenberger, O'Neill, & Csicsvari, 2020).

The studies discussed above have all been in rodents, but there is also an increasing amount of evidence for memory reactivation in humans. Recent experiments have made a lot of progress in this area, which will be discussed further in section 1.4.3. The studies reviewed here have given us a lot of information about what is presumed to be the neurophysiological basis of memory consolidation during sleep. Nevertheless, they are limited in the sense that they are usually not

causal – the relationship between spontaneous memory reactivation and memory consolidation is generally assessed using correlations between detected reactivation and improvements on memory tasks. A more causal link may be established using Targeted Memory Reactivation (TMR).

1.4.2 Targeted Memory Reactivation (TMR)

TMR is a technique whereby learning materials used in a task before sleep are coupled with sensory stimuli such as sounds or odours. During sleep, these sensory stimuli are covertly re-presented, to bias the brain towards memory reactivation of the cued materials. The goal is thus to selectively improve memory consolidation of the cued items and thereby increase performance on those items in subsequent memory tests. Already in the early days of memory reactivation research, there were some attempts at triggering these reactivations, for instance by using electrical shocks for fear conditioning in rats (Hars, Hennevin, & Pasques, 1985), auditory cues paired with Morse code learning (Guerrien, Dujardin, Mandai, Sockeel, & Leconte, 1989), and a clicking sound paired with a complex logic task (Smith & Weeden, 1990). However, these and similar studies had only few participants and often did not employ electrophysiological methods to check that participants were actually asleep (Oudiette & Paller, 2013). As such, the potential for cued memory reactivation during sleep remained relatively unexplored for a long time.

The revival of TMR came in 2007, when Rasch and colleagues published their study which used odour to improve declarative memory during sleep (Rasch et al., 2007). While participants learned object-location pairs on a computer screen, the scent of a rose was spread. Afterwards, as participants slept, this same scent was presented again during SWS without disturbing the participant. Those participants that smelled the scent during both learning and SWS showed better recall of the learned pairs compared to those who were not presented with the odour during either learning or SWS. Functional magnetic resonance imaging (fMRI) revealed activation of the hippocampus following odour presentation during SWS, but only when that same odour had been presented during learning. It has been hypothesised that this re-presentation of the rose scent reactivated the memory of the context of learning, which resulted in a benefit for the material learned in that context (Oudiette & Paller, 2013).

Increased specificity may be achieved using auditory cues rather than odours. This was first done in 2009, again using an object-location task (Rudoy, Voss, Westerberg, & Paller, 2009). Objects were paired with relevant sounds, e.g. a bark for a dog, a meow for a cat, etc. Half of these sounds were softly played to participants as they took a nap after learning, without disrupting their sleep. In a

memory test after the nap, recall of locations was better for the objects that had been cued during sleep compared to those that had not been cued. Thus, it was shown that TMR has the ability to improve very specific, even individual, memories.

TMR has since been applied in various sleep stages to strengthen various types of memory. By far the most popular stage for TMR has been SWS, and this has mainly been used to enhance performance on declarative memory tasks. As mentioned above, spatial location learning benefits from TMR during this stage (e.g. Diekelmann, Büchel, Born, & Rasch, 2011; Rasch et al., 2007; Rudoy et al., 2009), but also spatial navigation (Shimizu et al., 2018), word-picture associative learning (Cairney et al., 2018), foreign language learning (Göldi, van Poppel, Rasch, & Schreiner, 2019; Schreiner, Lehmann, & Rasch, 2015; Schreiner & Rasch, 2015), and several other types of declarative memories (Cairney, Sobczak, Lindsay, & Gaskell, 2017; Fuentemilla et al., 2013; Ritter, Strick, Bos, Van Baaren, & Dijksterhuis, 2012). Procedural skills also appear to benefit from SWS TMR, especially learning of a sequence of finger movements (Antony, Gobel, O'Hare, Reber, & Paller, 2012; Cousins, El-Deredy, Parkes, Hennies, & Lewis, 2014, 2016). This sequence aspect does not appear to be crucial for TMR to work, however, given that sensorimotor skill performance has also been shown to increase (Johnson, Scharf, Verceles, & Westlake, 2019; Johnson, Scharf, & Westlake, 2018). TMR during SWS (or more broadly, NREM sleep) has even been used to induce forgetting, by pairing an auditory tone with the act of forgetting and playing it again during sleep (Schechtman, Witkowski, Lampe, Wilson, & Paller, 2020; Simon, Gómez, & Nadel, 2018).

Successful TMR during Stage 2 sleep has mostly been associated with improvements in procedural skill (Laventure et al., 2016, 2018). In truth, because many studies have applied TMR during NREM more broadly, rather than specifically in S2 or SWS, distinguishing the effects of TMR during SWS and S2 can sometimes be difficult. Interestingly, however, S2 was the stage used in the only tactile TMR study, although no effects on procedural skill were found in this case (Pereira et al., 2017). A study where the name of a familiar snack item (e.g. 'skittles') was used as a TMR stimulus during S2 showed an increase in preference for that snack (Ai et al., 2018).

TMR during REM sleep has been attempted less frequently compared to NREM and with mixed results. Early studies into memory cueing, mentioned in the first paragraph of this section, mainly applied cues during REM sleep, but these studies suffered from methodological problems (reviewed in Oudiette & Paller, 2013). Several experiments using olfactory TMR during REM did not find any effects, either on declarative or procedural memory (Cordi, Diekelmann, Born, & Rasch, 2014; Laventure et al., 2016; Rasch et al., 2007). REM sleep has often been implicated in emotional

memory consolidation (Hutchison & Rathore, 2015), but TMR studies that have tested this have been largely unsuccessful (Lehmann, Schreiner, Seifritz, & Rasch, 2016; Rihm & Rasch, 2015). One study showed that auditory TMR during REM sleep enhanced accurate and false memories of faces, indicating an increase in integration or associative memory strength (Sterpenich et al., 2014).

TMR has also regularly been applied during wakefulness, often as a control for sleep TMR. This has not generally been shown to improve subsequent memory performance, though this may partly be due to the tasks participants completed as wake TMR took place. To prevent participants from rehearsing task material during wake TMR, they are often asked to do a demanding working memory task (e.g. Rudoy et al., 2009; Schreiner & Rasch, 2015). Wake TMR may be more successful when the accompanying task does not demand all of the participant's attention (Tambini, Berners-Lee, & Davachi, 2017; Tambini & Davachi, 2019).

Recently, a meta-analysis was conducted on 91 TMR studies (Hu, Cheng, Chiu, & Paller, 2020). Overall, TMR during SWS and Stage 2 was found to positively affect subsequent memory performance, whereas TMR during REM and wakefulness overall did not show an effect. Additionally, recall and reaction time tasks showed a bigger effect than recognition tests, and within-subject experiments showed a higher increase than those employing a between-subject design. There appeared to be no substantial difference between auditory and olfactory TMR in terms of effectiveness. Generally, TMR appears to be a fairly robust technique, especially when applied in SWS and Stage 2 sleep. At the moment, very few REM TMR studies have been carried out, which means that the role of REM sleep in memory consolidation remains somewhat of a mystery.

Importantly, Bendor and Wilson (2012) have shown in rats that TMR cues can indeed bias memory reactivation towards a reactivation of the memory associated with that cue. A major benefit of TMR is therefore that it provides a fairly narrow time window during which a supposed memory reactivation is to take place. Thus, TMR has paved the way for significant advances in the detection and characterisation of human (sleep) reactivations.

1.4.3 Detection of human (sleep) reactivations

It is rather more difficult to detect memory reactivations in humans compared to animals, because researchers primarily (though not exclusively) have to rely on imaging methods rather than implanted electrodes. Some of the first evidence of memory reactivation during sleep in humans

came from studies using positron-emission tomography and fMRI. These studies looked at blood flow activity in specific brain areas to detect spontaneous memory reactivation, particularly during REM sleep. Using these methods, they were able to show that brain areas that were active during a task participants completed while awake were significantly more active during sleep in participants that completed the task compared to those that did not (Maquet et al., 2000). In REM sleep, this reactivation was related to the sequential structure of a procedural task, which participants learned implicitly, rather than basic visuomotor properties (Peigneux et al., 2003). Using a completely different task, this same group was able to show that in SWS, reactivation correlated positively with performance on the retest of spatial memory the next day (Peigneux et al., 2004). A later fMRI study also found that reactivation of a non-hippocampal visual perception task during NREM sleep was positively correlated with subsequent improvement on the task (Yotsumoto et al., 2009).

Although these studies indicated that reactivation likely took place, a key aspect of demonstrating memory reactivation is the correspondence between the pattern of brain activation during learning and during the presumed reactivation period. The studies mentioned above could not address that question, but methods such as multivariate pattern analysis (MVPA) (Haxby et al., 2001) and representational similarity analysis (RSA) (Kriegeskorte, 2008) may be able to. Schönauer and colleagues were the first who used MVPA to classify, during sleep, whether participants had viewed images of faces or houses during a wake learning session (2017). They were able to detect learning-related patterns of EEG activity during both NREM and REM sleep, but only the strength of classification during SWS was correlated with memory performance after waking up.

By combining TMR and RSA, Cairney and colleagues (2018) showed that NREM cueing of an associative (word-picture) task increased fast spindle activity, and during this activity they were able to tell which category the picture that had been associated with the played word belonged to. Moreover, the distinctiveness of these categories as decoded by the RSA was positively associated with the next-day memory benefit for cued over uncued items. Similarly, in an fMRI study using odour TMR of different categories during NREM sleep, MVPA was able to distinguish those different categories (Shanahan, Gjorgieva, Paller, Kahnt, & Gottfried, 2018). Odours produced category-specific reactivation in the ventromedial prefrontal cortex, and the degree to which it did so was related to the cueing benefit for the specific categories. Combined, these studies indicate that TMR indeed triggers behaviourally relevant memory reactivations during sleep.

Notably, these three studies made use of classifiers that were both trained and tested on sleep data. As mentioned above, to demonstrate reactivation of learning-related brain activity, it is essential that the activity during learning is taken into account. This presents a significant challenge, because oscillatory activity during sleep is so different from that of wakefulness. Nevertheless, two studies to date have accomplished this. Schreiner and colleagues (2018) used a vocabulary learning paradigm with TMR. Cues presented during NREM sleep prompted a neural pattern that was similar to that of processing during the task, and this seemed to be driven by theta oscillations in both instances. In contrast, Belal and colleagues (2018) used a multivariate pattern approach to classify memory reactivation of a procedural task. A classifier was trained on EEG data from a motor imagery task during which participants imagined previously learned finger movements, and tested on the EEG after TMR during NREM sleep. Memory reactivation was detected in SWS in all participants, and in five out of fourteen participants during S2.

Finally, one study has probed memory reactivation using intracranial EEG (iEEG) in human epilepsy patients, a method that opens up the possibility of looking at hippocampal sharp wave-ripples. Using RSA, stimulus-specific activity from encoding was detected to spontaneously re-occur both during waking rest and sleep (Zhang et al., 2018). Only reactivations that had been triggered by ripples during NREM sleep were related to memory consolidation.

Next to classifiers, electrophysiological data computed from the sleep EEG such as event-related potentials (ERPs) and time-frequency analyses are sometimes used as an indication of memory reactivation. Generally, this is done by comparing the electrophysiological activity resulting from cues paired with items that were later remembered and later forgotten, or by comparing such activity resulting from memory-related cues (TMR cues) to activity resulting from other sounds (control cues). Differences between these are thought to arise from the memory content that accompanies successful TMR cues.

Although there have not been many studies which have examined the electrophysiological effects of TMR in such a way, some general findings have emerged. In a series of vocabulary learning experiments, Schreiner and colleagues compared ERPs and time-frequency data resulting from words that were later remembered to words that were later forgotten. They found that successful reactivation during NREM sleep was accompanied by an increase in theta and spindle power approximately 500-1000 ms after cue onset, and a large frontal negativity in ERPs about 800-1100 ms after the cue (Lehmann et al., 2016; Schreiner et al., 2015; Schreiner & Rasch, 2015). An increase in spindle power after TMR cues during NREM sleep was also found by Cairney and colleagues, in

this case in comparison to control cues that had not been heard before (Cairney et al., 2018). In contrast to the Schreiner studies, here the increase in spindle power was found approximately 2 seconds after the cue.

The detection of memory reactivation in humans has almost exclusively taken place during NREM sleep. Some of the early work (Maquet et al., 2000; Peigneux et al., 2003) and one more recent paper (Schönauer et al., 2017) do suggest that reactivation also takes place during human REM sleep, but little is known about its mechanisms, relationship to NREM reactivation, and behavioural effects. Some evidence suggests that reactivation in this stage may be related to more complicated functions of sleep-related memory consolidation, such as the restructuring of memories during sleep (Sterpenich et al., 2014), but this has only been sparingly explored.

1.5 The restructuring of memories during sleep

An intriguing aspect of sleep-dependent memory consolidation is that it not only strengthens memories or protects them from interference, but also qualitatively changes memories. Here, I will discuss three ways in which this may present itself, corresponding to the three types of experiments which form Chapters 2-4 of this thesis.

1.5.1 Regularity abstraction and generalisation

We have previously seen that sleep (with or without TMR) can improve performance on a range of declarative and procedural tasks. This alone is not sufficient evidence in favour of the idea that sleep is involved in the reorganisation of memory representations, since these performance improvements may have arisen through the consolidation of connections that were established during learning. However, there are several studies which have demonstrated that sleep aids the abstraction of regularities in learned materials, and supports processes whereby new memories are integrated with previously learned information (Rasch & Born, 2013).

A striking example of this function comes from studies looking at the influence of sleep on statistical learning. In one paradigm, participants heard a series of tones which followed a sequential structure that was probabilistically determined (Durrant, Cairney, & Lewis, 2013, 2016; Durrant, Taylor, Cairney, & Lewis, 2011). After exposure to these tones for a number of minutes, participants completed a short test and either slept or stayed awake. It was found that sleep enhanced the recognition of new sequences that followed the same statistical pattern significantly more than

wakefulness, and these effects were related to SWS. Another paradigm looking at statistical learning is the weather prediction task, which involves the presentation of abstract images that probabilistically predict either sun or rain using a complex rule comprising combinations of images. Sleep improved performance on this task, and this improvement was correlated to REM sleep (Barsky, Tucker, & Stickgold, 2015; Djonlagic et al., 2009). Furthermore, sleep increased participants' ability to evaluate the probabilities of predicting sun or rain for single images.

Such regularity abstraction may also be thought of as schema formation. A schema is conceptualised as a framework of knowledge, which, once established, generalises to new information and thereby accelerates the pace at which this new information can be consolidated (Tse et al., 2007). Sleep may further expedite this process, as shown for example in work looking at the generalisation of categorical learning. In such experiments, participants learn about stimuli which belong to different categories. Then, they are tested to see whether they are able to generalise the category knowledge by attempting to classify new stimuli or the category prototypes which had never been seen. Several studies have found that sleep benefits this category generalisation, especially in children (Friedrich, Mölle, Friederici, & Born, 2019; Friedrich, Wilhelm, Born, & Friederici, 2015; Friedrich, Wilhelm, Mölle, Born, & Friederici, 2017; Graveline & Wamsley, 2017; Sandoval, Leclerc, & Gómez, 2017), though others have found no sleep effect (Maddox et al., 2011; Werchan & Gómez, 2014).

Additional evidence comes from perceptual and grammar learning, such as a study where training of synthetic speech perception generalised to words that had not been heard before (Fenn, Nusbaum, & Margoliash, 2003). Similarly, both children and adults are able to detect hidden grammatical rules in an artificial language and generalise them to new sentences, and this effect appears to be related to SWS (Batterink & Paller, 2017; Gaskell et al., 2014; Gómez, Bootzin, & Nadel, 2006; Simon et al., 2017).

In Chapter 2 of this thesis, I have employed a paradigm that is often used to examine sleep's role in pattern extraction: the serial reaction time task (SRTT). In this task, participants are asked to respond quickly and accurately to a series of cues on the screen by pressing the corresponding buttons. Unknown to the participants, the cues are presented in a sequence, and by learning this sequence (either implicitly or explicitly) participants can greatly reduce their reaction times. Sleep has often been shown to improve procedural skill as measured by a decrease in reaction times (Cousins et al., 2014, 2016; Maquet et al., 2000), and explicit knowledge of the sequence also increases in some studies (Cousins et al., 2014; Diekelmann, Born, & Rasch, 2016; Wilhelm et al.,

2013). Other evidence for memory generalisation in procedural tasks includes the fact that learning a sequence in one hand followed by sleep leads to improvements on the original (but not the mirror) sequence in both the trained and the untrained hand (Cohen, Pascual-Leone, Press, & Robertson, 2005; Witt, Margraf, Bieber, Born, & Deuschl, 2010). This suggests that sleep consolidates the representation of the sequence independently from the required movements. Moreover, motor sequence learning can be achieved by observation (Van Der Werf, Van Der Helm, Schoonheim, Ridderikhoff, & Van Someren, 2009) and motor imagery (Debarnot, Creveaux, Collet, Doyon, & Guillot, 2009), when these are followed by sleep.

A key question is *how* sleep aids regularity abstraction and generalisation. The Active System Consolidation view, discussed in section 1.3.2, maintains that the process of repeated memory reactivation drives consolidation. During the process of reorganisation of memory into long-term stores, regularities between memories are extracted (Born & Wilhelm, 2012; McClelland et al., 1995). The Synaptic Homeostasis Hypothesis, on the other hand, is built around the idea that SWS leads to large-scale synaptic downscaling, which is part of homeostatic processes that renormalize synaptic strength after potentiation during wake (Tononi & Cirelli, 2003). During this downscaling, the shared features of memories are maintained, while other, less relevant features are decreased in strength, thereby extracting regularities between memories (Nere, Hashmi, Cirelli, & Tononi, 2013). These two processes may also work together, as in the 'information overlap to abstract' (iOtA) model (Lewis & Durrant, 2011). In this model, linked memories are strengthened by memory reactivation and followed by general synaptic downscaling. Together, these processes ensure that overlapping elements are maintained and others erased, leading to schemas that represent the relationships between memories.

1.5.2 Associative inference

A slightly different angle for examining sleep-mediated memory restructuring comes from experiments that deal with associative inference, which is also sometimes called relational memory or learning. In such experiments, participants learn the relationships between pairs of images, for instance faces and objects or scenes. Participants explicitly learn pairs of type A-B, and pairs of type B-C. In other words, the learned pairs contain a shared element (B), and through this the relationship between A and C can be found. Crucially, this 'remote' relationship was never learned during training, but must be inferred. Studies using this type of task during wakefulness (without the involvement of sleep) have characterised relational learning as hippocampus-dependent (Bunsey & Eichenbaum, 1996; Preston, Shrager, Dudukovic, & Gabrieli, 2004; Schlichting, Guarino,

Schapiro, Turk-Browne, & Preston, 2017), and the ventromedial prefrontal cortex appears to be essential in the ability to make inferences (Spalding et al., 2018; Zeithamova, Dominick, & Preston, 2012).

Two studies have used this type of task to investigate the relationship between sleep and associative inference. Lau and colleagues (2010) found that both direct and remote associations benefitted from a nap as compared to an equal period of wakefulness. Relational memory (inference) performance correlated with the amount of SWS obtained during the nap, but no relationship between sleep stages and memory for the learned associations was found. Notably, participants who entered REM sleep during the nap were excluded from the analyses. This is important, because a similar nap study found a positive relationship between the percentage of REM sleep obtained and performance on the inference test (Alger & Payne, 2016). Like Lau and colleagues (2010), this study also found performance improvements after sleep compared to wake on both direct and remote associations. Interestingly, no studies to date have used TMR to illuminate the role of sleep in associative inference.

A slightly different but related paradigm is that of transitive inference. During the task, participants had to choose between two abstract images which, unbeknownst to them, were organised by a hierarchy of preferences: A > B > C > D > E > F. Participants always chose between adjacent pairs (e.g. A>B, B>C, E>F) during training, and through trial and error they had to discover the underlying hierarchy. After a period of time they were tested on the learned pairs, and on unlearned inference pairs (B>D, C>E, and B>E). Participants showed a higher accuracy on the inference pairs after sleep compared to wake, but this did not seem to be paired with an increase in awareness of the hierarchy (Ellenbogen, Hu, Payne, Titone, & Walker, 2007). This experiment was subsequently replicated in another study, which showed that the process of trial and error during training was crucial for later successful transitive inference (Werchan & Gómez, 2013).

1.5.3 Creativity

Creativity is another type of skill that may benefit from the reprocessing and restructuring of memory during sleep. Sleep has been said to inspire creative ideas, at least anecdotally. Nevertheless, the link between creativity and memory restructuring (during sleep) may not be obvious for many.

The standard definition of creativity consists of two parts: originality and appropriateness (Runco & Jaeger, 2012; Stein, 1953; Sternberg & Lubart, 1996). Originality – also called novelty, uniqueness, unusualness – is generally seen as the core concept of creativity. Something that is not original is not creative, but this originality is not sufficient. Mere originality may arise through randomness and may be generated by monkeys on type writers, and these random but original ideas or products may lack any application. In other words, there may be a good reason they are unique or unusual. Thus, to be creative, something original must also be appropriate. This appropriateness may be called effectiveness, usefulness, utility, fit, or even value, but researchers broadly agree that it is an important aspect of creativity (Runco & Jaeger, 2012).

In thinking through this definition, the important role of memory in creativity already becomes apparent. To come up with something original and appropriate, one must remember what came before, what meaning or properties things have, and how all these elements fit together. In other words, creative thinking involves combining aspects of memory that may not have previously been combined, forming associations, and through this process creating something that is useful but also contains novel aspects.

The idea that creativity is an associative memory process has long been proposed. Mednick (1962) suggested that associative abilities are the key to creativity, to the extent that variability in the way concepts are retrieved or associated with each other in people's semantic networks leads to individual differences in creativity. Indeed, associative abilities have been shown to affect creative performance (Benedek, Könen, & Neubauer, 2012). This idea is also supported by studies on the semantic memory network of low and high creative participants, which found that those who score low on creativity have an associative network that is less connected, more spread out, and contains more sub-parts (Benedek et al., 2017; Bernard, Kenett, Ovando-Tellez, Benedek, & Volle, 2019; Kenett, Anaki, & Faust, 2014; Kenett & Faust, 2019).

Creativity is further thought to be supported by executive processes, whereby top-down control allows for strategic memory retrieval and knowledge manipulation (Beaty & Silvia, 2012). For instance, one study found that low and high creative participants did not differ in their associative hierarchies, and instead suggested that high creative people more effectively access their associative memory (Benedek & Neubauer, 2013). Crucially, semantic memory plays a key role in both the associative and executive accounts of creativity (Abraham & Bubic, 2015), and it is likely that both processes work together to generate ideas and satisfy task goals (Beaty, Benedek, Silvia, & Schacter, 2016).

As a final thought on the link between memory and creativity during wakefulness, it has also been suggested that memory may impede creativity when prior knowledge or old ideas lead to mental fixation (Storm, Ditta, & George, 2020). A classic example is the two-string problem, where participants have to join two pieces of string hanging from a ceiling (Maier, 1931). They cannot reach both strings at the same time, but they can use the objects in the room, one of which is a pair of pliers. Participants could use these pliers as a pendulum to swing one piece of string towards them while holding the other. However, many participants do not reach this solution; likely because they were more inclined to think of the pliers' common function. As such, (temporary) forgetting may actually be beneficial for creativity. It has been suggested that the positive creative effects of incubation – a period away from a problem or task – may be partially due to forgetting during this interval (Kohn & Smith, 2009; Koppel & Storm, 2014; Storm & Angello, 2010).

How, then, does creativity fit in with theories of sleep as a process of memory restructuring? In section 1.5.1, we explored the role of sleep in regularity abstraction, which may in turn benefit creativity. When memories are stored in gist-based schemas, this can allow for the analogical transfer of one experience to another (Gick & Holyoak, 1983). By extracting regularities, we see the similarities rather than the differences, and this allows us to flexibly make use of prior experiences when generating creative solutions or ideas. Sleep has indeed been shown to improve problem solving by supporting analogical transfer (Monaghan et al., 2015). Additionally, sleep may strengthen associative processes that promote creativity. The associative inference task mentioned in section 1.5.2 involves making an inferential connection between two stimuli that were not learned together but were remotely associated, and this task has been known to benefit from sleep (Alger & Payne, 2016; Lau et al., 2010).

The Broader form of the iOtA model (BiOtA) has explicitly formulated hypotheses about the role of sleep in boosting creative problem solving (Lewis et al., 2018). It suggests that memory reactivation during NREM sleep promotes the abstraction of gist. In turn, REM sleep is considered to be a period of high cortical connectivity, during which overlapping reactivation may enhance the formation of novel associations. The interleaving of these two stages may consequently stimulate analogical problem solving.

Of additional relevance are two theories that relate to semantic memory restructuring, both of which have been investigated using the Deese-Roediger-McDermott (DRM) paradigm. In this paradigm, participants have to learn lists of semantically related words which lack the critical word that summarizes each list. Sleep has been shown to increase false memories for the critical word

(Diekelmann, Born, & Wagner, 2010; McKeon, Pace-Schott, & Spencer, 2012; Monaghan, Shaw, Ashworth-Lord, & Newbury, 2017; Pardilla-Delgado & Payne, 2017; Payne et al., 2009), depending on the characteristics of the memory task used (Newbury & Monaghan, 2019). These results have been explained as another example of the role of sleep in gist abstraction, because the critical word represents the gist of a word list (e.g. Diekelmann et al., 2010). However, another hypothesis has proposed that these false memories are caused by spreading activation between associated concepts in the semantic network. Sleep, then, is thought to further boost this spreading activation (e.g. Monaghan et al., 2017). Notably, sleep may affect creativity through both of these processes.

Creativity, like memory, is a construct that consists of many different aspects, and thus it may be the case that different aspects of creativity may be differently related to memory and sleep. A major distinction that is often made is between convergent and divergent thinking (Guilford, 1967). Divergent thinking involves the generation of many different ideas, and is exemplified by the Alternative Uses Task (AUT) (Guilford, 1967). In this task, participants must come up with unusual uses for common household objects (e.g. a tin could be used as a jar, a cookie cutter, or a musical instrument). Convergent thinking, on the other hand, is seen as a process of generating one possible (correct) outcome. For example, in the Remote Associates Task (RAT) (Mednick, 1962), participants must find a common concept that unites three given words: e.g. 'dream', 'break', 'light' (solution: day). Divergent thinking has generally been more closely associated with creativity, since it explicitly involves the generation of novel ideas, but it has been argued that the evaluation of these ideas (via convergent thinking) is crucial for creativity (Cropley, 2006).

Studies of the effect of sleep on creativity have thus far mainly looked at convergent tasks. Two experiments have used the RAT, with both finding benefits of sleep. One showed that improvements were related to REM sleep (Cai, Mednick, Harrison, Kanady, & Mednick, 2009), and the other found an effect specifically for more difficult problems (Sio, Monaghan, & Ormerod, 2013). Interestingly, both studies interpreted their findings through the lens of spreading activation. On the topic of abstraction, one study found that REM sleep awakenings were more beneficial for anagram problem solving than NREM awakenings (Walker, Liston, Hobson, & Stickgold, 2002). Another showed that sleep more often led to insight into the hidden rule that governed a task than wake (Wagner, Gais, Haider, Verleger, & Born, 2004). Sleep has also been found to benefit video game problem solving (Beijamini, Pereira, Cini, & Louzada, 2014), and puzzle solving (Sanders, Osburn, Paller, & Beeman, 2019), and both effects appeared to be due to SWS. Only one study has used a divergent task. Participants were given a problem for which they had to find a creative

solution, and odour TMR during an entire night of sleep made participants more creative and more able to choose their most creative solution (Ritter et al., 2012).

While these are all success stories, it is important to note that there are also studies that have found no effect of sleep on creativity. Schönauer and colleagues (2018), for example, showed that a nap did not improve the solving of magic tricks and classical insight problems compared to an equal time spent awake. Similarly, a nap did not lead to an increase in riddle solving compared to wake (Brodt, Pöhlchen, Täumer, Gais, & Schönauer, 2018). Recently, another experiment using a murder mystery video game also found no beneficial effects of sleep on the quality and creativity of the solutions participants proposed (Hołda, Głodek, Dankiewicz-Berger, Skrzypińska, & Szmigielska, 2020). Given the disparate findings and the lack of divergent thinking tasks used, the role of sleep in creativity remains unclear. Creativity can be seen as a symbol of mental reorganisation, and as such it would be especially interesting to determine to what extent and in what way creativity depends upon the reprocessing and restructuring of memory during sleep.

1.6 Summary

Sleep consists of several different stages, most notably slow-wave sleep (SWS) and rapid eye-movement (REM) sleep. We spend about 1/3 of our lives asleep, but we still do not know exactly why. One of the proposed functions that has received a lot of attention, and which is the focus of this thesis, is the contribution of sleep to memory. Specifically, sleep has been thought to play an active role in the consolidation of memory: the process of stabilising memories after they have been encoded. This role is thought to be achieved through memory reactivation. The reprocessing of a memory trace occurs spontaneously during wake and sleep, and this phenomenon has been associated with improvements in memory. The reactivation of memory may also be triggered or biased using a technique called targeted memory reactivation (TMR). This involves pairing elements in a task with sensory stimuli during learning, and re-presenting these sensory stimuli during sleep to prompt the memory of the learned materials. By using this technique, the detection of memory reactivation in humans has achieved significant advances, although detection of reactivation during REM sleep in humans remains somewhat elusive.

Sleep, particularly SWS, has been shown to strengthen and restructure a variety of memory types. The focus of this thesis is on restructuring; on the idea that memory is not only enhanced but also transformed by sleep. This may take the form of regularity abstraction and integration, associative

inference, and creativity. Regularity abstraction, also called gist abstraction, has received a substantial amount of attention and may be promoted by SWS, REM, or their combination. It is thought to be achieved by repeated memory reactivation of overlapping elements of our experiences, leading to the extraction of regularities between them. Associative inference involves making novel connections between items that were not learned together but whose relationship can be inferred. This type of relational learning has been associated with both REM and SWS. Creativity, finally, is the creation of something original and useful. It has strong ties to memory, and thus may be enhanced by processes of memory restructuring during sleep. However, this link still needs to be explored more, to clarify which creative tasks benefit from sleep under which circumstances.

1.7 Research objectives

This thesis consists of three experimental chapters that aim to explore the role of sleep in the reprocessing and restructuring of memory. **Chapter 2** examines this with the serial reaction time task, a task in which performance can be improved by both the implicit and explicit abstraction of the sequences used. We performed TMR in SWS and REM to detect memory reactivation and behavioural effects related to these sleep stages. We were further interested in whether such effects would be found across the board or in one hand specifically, which may help to interpret some of the mechanisms or priorities of sleep-dependent memory restructuring. In **Chapter 3**, we explore associative inference, again using TMR in SWS and REM. Participants learned associations between faces and scenes, where each scene was related to two faces. We wanted to see whether participants' ability to infer the relationship between two remotely associated faces would be enhanced by TMR in either sleep stage. This would indicate that those sleep stages might be involved in making novel connections. Finally, in **Chapter 4**, we look at the effects of sleep and wakefulness on two creative tasks. Participants took part in three sessions, each twelve hours apart, starting either in the morning or in the evening. We chose to use one task that is highly dependent upon a participant's semantic network, to examine whether an interval filled with sleep would lead to more distant associations. Furthermore, we chose a classic divergent thinking task, because the relationship between sleep and divergent thinking has thus far not been sufficiently evaluated.

CHAPTER 2

Targeted memory reactivation of a serial reaction time task in SWS, but not REM, preferentially benefits the non-dominant hand

This chapter is based on a manuscript which is being prepared for submission for peer review. The classification analysis was developed and carried out by Mahmoud E. A. Abdellahi. The 'Classification' sections in the Methods and Results were written jointly with him.

2.1 Abstract

Targeted memory reactivation (TMR) is a technique by which sounds paired with learned information can be used to cue neural reactivation of that information during sleep. While TMR in slow-wave sleep (SWS) has been shown to strengthen procedural memories, it is unclear whether TMR in rapid eye movement (REM) sleep, a state strongly associated with motor consolidation, provides equivalent benefit. Furthermore, it is unclear whether this technique influences dominant and non-dominant hands equally. We used TMR of a two-handed serial reaction time task (SRTT) during both SWS and REM in thirty-two human adults (sixteen female) to examine how stimulation in each sleep stage impacts on dominant (right) and non-dominant hands. Additionally, we developed a machine learning classifier to detect memory reactivation in sleep using scalp electroencephalography. Interestingly, the TMR related performance improvement occurred after cueing in SWS, but not REM, and was present in the non-dominant but not the dominant hand. Furthermore, our classifier reliably detected memory reactivation during SWS but not REM. Nevertheless, event-related potentials to left- (non-dominant) and right-handed cues differed significantly in REM, but not SWS. These results show that TMR is more effective in the non-dominant hand. Furthermore, while the brain processes TMR during both SWS and REM, such cueing only leads to behavioural benefit when applied in SWS. Reactivation may take place during REM, but its link with memory consolidation remains unclear.

2.2 Introduction

Memories consolidate across sleep (Diekelmann & Born, 2010; Rasch & Born, 2013), and this is facilitated by offline reactivation in which task related brain activity is reinstated during sleep (Skaggs & McNaughton, 1996; Wilson & McNaughton, 1994). Targeted memory reactivation (TMR) can be used to influence memory consolidation by biasing memory reactivation, for instance by playing sounds that were previously linked to items learned in wake during subsequent sleep. TMR has been shown to trigger memory reactivation (Belal et al., 2018; Schreiner et al., 2018; Shanahan et al., 2018), and to influence behaviour after sleep, for example by improving episodic (Cellini & Cappuzo, 2018; Rasch et al., 2007; Rudoy et al., 2009) and procedural skill consolidation (Antony et al., 2012; Schönauer et al., 2014). In this study, we set out to determine whether TMR equally impacts on motor memories in the dominant and non-dominant hand. This novel experimental design will give us more information about the specificity of memory consolidation during sleep. Will performance in each hand benefit from TMR in a similar matter, or will TMR improve one hand

more than the other? Put differently: which memories will be strengthened through TMR, and which do not benefit as much? Once we have a clearer idea about the (types of) memories that are selectively improved, we can start looking into how the brain makes this selection. We were also interested to determine whether the sleep stage in which TMR was applied would determine the extent to which it elicited reactivation and impacted on consolidation. In other words, which sleep stage should TMR be applied in to be most beneficial for this task? This has also not been examined before. We used the highly sleep dependent serial reaction time task (SRTT), which has already been shown to be sensitive to TMR in SWS (Cousins et al., 2014).

To better determine the relationship between TMR and reactivation, as well as between reactivation and behavioural consolidation, we developed an EEG classifier to detect reactivation using scalp EEG. Our work follows prior studies that have identified reactivation using EEG (Belal et al., 2018; Cairney et al., 2018; Schreiner et al., 2018), intracranial recordings (Zhang et al., 2018), and fMRI (Deuker et al., 2013; Shanahan et al., 2018).

Studies of sleep dependent consolidation in motor skills tend to focus on the non-dominant hand (Korman et al., 2007; Korman, Raz, Flash, & Karni, 2003; Spencer, Sunm, & Ivry, 2006; Walker, Brakefield, Hobson, & Stickgold, 2003; Walker, Brakefield, Morgan, Hobson, & Stickgold, 2002; Walker, Brakefield, Seidman, et al., 2003; Walker, Stickgold, Alsop, Gaab, & Schlaug, 2005). This is also true for TMR studies of procedural memory (Antony et al., 2012; Cousins et al., 2014, 2016; Schönauer et al., 2014). The non-dominant hand is typically chosen to reduce the influence of pre-existing motor skills (Maquet, Schwartz, Passingham, & Frith, 2003) and because greater performance gains may be possible in this hand (Ridding & Flavel, 2006; Spencer et al., 2006). However, we are unaware of any study comparing the impact of sleep dependent memory consolidation on dominant and non-dominant hands. As mentioned above, this would give us a lot of information about the specificity of memory consolidation. We therefore set out to examine this using TMR, as well as our EEG classifier, which could distinguish between reactivation relating to the dominant and non-dominant hand. Note that because we used exclusively right-handed participants, like is common in the literature, the dominant hand is always the right hand and the non-dominant hand is always the left hand. These terms are thus used interchangeably throughout this chapter, and the relevance of their difference is explored in the discussion.

Slow-wave sleep (SWS) is important for memory consolidation, and TMR is often applied during non-REM stages such as SWS or Stage 2 (e.g. Antony et al., 2018; Cairney et al., 2014; Fuentemilla et al., 2013; Hauner et al., 2013; Rasch et al., 2007; Rudoy et al., 2009). Nevertheless, early studies

show that tasks with a procedural memory component benefit from rapid eye movement (REM) sleep (Karni, Tanne, Rubenstein, Askenasy, & Sagi, 1994; Smith, 1993, 1995, 2001; Smith & Smith, 2003) and brain regions involved in the serial reaction time task (SRTT) are reactivated during REM (Maquet et al., 2000; Peigneux et al., 2003). Furthermore, rodent studies have clearly identified replay in REM (Booth & Poe, 2006; Howe et al., 2019; Louie & Wilson, 2001; Poe et al., 2000), while human work has suggested that REM replay may be stronger than replay in other sleep stages (Schönauer et al., 2017). Finally, prior work on this task in our lab showed that the impacts of SWS TMR were mediated by subsequent REM, which was associated with widespread plasticity in the motor system (Cousins et al., 2016). Nevertheless, the influence of REM TMR on this task has never been examined. Given the literature, we set out to examine the impacts of auditory TMR in REM on the SRTT, and to determine how these compared to known benefits associated with TMR in SWS.

First, we expected to replicate the finding that TMR during SWS would improve behaviour on the cued sequence as compared to the uncued sequence, as shown by several previous studies (most notably Cousins et al., 2014, 2016). With regards to REM TMR, we hypothesised that this would also improve behaviour on the cued sequence compared to the uncued sequence, given the fact that REM sleep has been implicated in this task. Following from the fact that most procedural TMR studies use the non-dominant hand only, we expected that this hand would particularly benefit from TMR, compared to the dominant hand.

In terms of the classifier results, we expected to be able to classify TMR during SWS, as this has been shown to be possible (Belal et al., 2018). In this context, classification is considered possible if its accuracy is significantly higher than chance. Wake-to-sleep classification of REM TMR has not been shown before, and furthermore the effect of REM TMR on this task is still unclear. Therefore, although we expected that classification of TMR during REM sleep would be possible in principle, we were unsure that our classification method would work in practice, making this analysis exploratory. We additionally hypothesised, based on the idea that electrophysiological data can indicate memory reactivation (Cairney et al., 2018; Lehmann et al., 2016; Schreiner et al., 2015; Schreiner & Rasch, 2015), that we would find differences between cues relating to the dominant and non-dominant hand in our analyses of event-related potentials and time-frequency data, in both the REM and the SWS group. Finally, based on the study by Cousins and colleagues (2014), we expected the cueing benefit in the SWS group to correlate with spindle laterality.

2.3 Materials and Methods

2.3.1 Participants

Thirty-five healthy right-handed, non-smoking participants were recruited for this study and randomly assigned to either the SWS- or REM-group, with the constraint of gender balance. Three participants were excluded, due to spending <30 minutes in SWS during the experimental night (*n* = 1), experimenter error resulting in TMR during wake (*n* = 1), and no evidence of learning during the pre-sleep training (*n* = 1). Sixteen participants remained in the REM Cued group (8 female, mean age 23.6 years) and sixteen in the SWS Cued group (8 female, mean age 23.1 years).

This sample size was based upon a previous study using the same task and TMR during SWS (Cousins et al., 2014). We conducted a post-hoc power analysis to examine the achieved power in our result of interest: the overnight behavioural effect of SWS TMR in the left hand. We used G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009), a one-sided Wilcoxon signed-rank test (matched pairs), and an α of 0.05. The effect size dz of 0.759 was calculated based on the means (56.828 vs 24.715), standard deviations (57.350 vs 48.137), and correlation (0.691) of the groups (cued vs uncued, respectively). This showed that the final sample size of thirteen participants (after additional exclusions) used in this behavioural analysis led to an achieved power of approximately 80.7%. In other words, we achieved a reasonable amount of power for the large effect we found, but the amount of participants would be insufficient to detect smaller effects.

All participants had normal or corrected-to-normal vision, normal hearing, and no history of physical, psychological, neurological, or sleep disorders. Responses in a pre-screening questionnaire reported no stressful life events, a generally regular sleep-wake rhythm in the month before the study, and no regular night work or cross-continental travel in the two months before the study. Participants were not taking any psychologically active medication or substances and agreed to abstain from alcohol and caffeine in the 24 and 12 hours prior to the start of the study, respectively. Subjects also agreed not to nap or participate in extreme physical exercise during the experiment. This study was approved by the School of Psychology, Cardiff University Research Ethics Committee, and all participants gave written informed consent.

2.3.2 Experimental tasks and design

Design

The study consisted of an adaptation night, which allowed participants to get used to sleeping in the lab with electrodes, and an experimental night during which they performed the behavioural tasks. The study design is outlined in Figure 2.1. There was at least one but no more than three nights between the adaptation and experimental nights.

Upon arrival for the adaptation night, around 9:30pm, participants changed into their sleepwear. At this time, their ability to use internal visual and kinaesthetic imagery was measured with a shortened version of the Movement Imagery Questionnaire-3 (MIQ-3; Williams et al., 2012) and handedness was assessed with a short version of the Edinburgh Handedness Inventory (EHI; Veale, 2014). They were then fitted for polysomnography (PSG) recording. Subjects' alertness was assessed by the Karolinska Sleepiness Scale (KSS; Åkerstedt & Gillberg, 1990) and the Stanford Sleepiness Scale (SSS; Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973) before going to bed around 11-11:30pm. Since our effect of interest was the within-participant effect of TMR, and not any between-participant effects of sleepiness, the KSS and SSS were not analysed further.

During the adaptation night, we played the tones that the participants would later (during the experimental training) learn to associate with one of the learned sequences. As these tones had not yet been associated with a memory, playing them during the adaptation night meant that they could be used as a control for analyses looking at the neural signature of memory reactivation. After 7-8 hours of sleep, participants were awakened. They then rated the sleep quality of the night with an adapted and translated version of a German sleep quality questionnaire (SQQ; Görtelmeyer, 1985). This questionnaire allowed us to compare the sleep quality between the adaptation and experimental night, and also included questions to probe whether participants had heard any of the TMR sounds during the night. Finally, subjects completed the KSS and SSS again. After removing the electrodes participants were offered the opportunity to shower before leaving the lab.

On the experimental night, participants arrived around 7:30pm and then changed into their sleepwear. They completed the Pittsburgh Sleep Quality Index (PSQI; Buysse, Reynolds, Monk, Berman, & Kupfer, 1989) to report their sleep quality over the past month. Participants' answers on the PSQI were used to confirm that they had slept well and regularly in the past month. As expected, because participants had already been pre-selected on being self-reported regular sleepers, no participants had to be excluded based on the outcome of the PSQI. Participants were subsequently fitted for PSG recording, after which they performed the serial reaction time task (SRTT; 50-60 minutes) and the motor imagery task (IMG; 30 minutes). Before each task, the KSS and SSS were completed to measure alertness. Participants were ready for bed around 11pm. During

the night, the tones that had been played during the adaptation night, and again during one of the learned sequences, were replayed. The sleep stage of cueing was the same in experimental and adaptation nights (SWS or REM, depending on the group the participant was in). As these tones were now associated with the SRTT, we expected that cueing them would trigger reactivation of this association (a memory) (Belal et al., 2018; Cousins et al., 2014, 2016). After 7-8 hours of sleep, participants were awakened and allowed at least 20 minutes to overcome sleep inertia. During this time, participants were given the opportunity to eat and drink something before completing the sleep quality questionnaire. Participants then completed the same tasks again in reverse order (IMG first, SRTT second), each preceded by the KSS and SSS. Finally, participants completed an explicit sequence memory test, by marking the sequence order on a printout containing pictures of the (empty) screen arranged vertically (Figure 2.1e). This was done for both sequences; the order was counterbalanced across participants. Tasks were presented on a computer screen with resolution 1024 x 768 pixels and using Matlab 6.5 (The MathWorks Inc., Natick, MA, 2000) and Cogent 2000 (Functional Imaging Laboratory, Institute for Cognitive Neuroscience, University College London). The tones were played through noise-cancelling headphones (Sony MDR-ZX110NA) during the tasks and through speakers (Dell A225) during sleep.

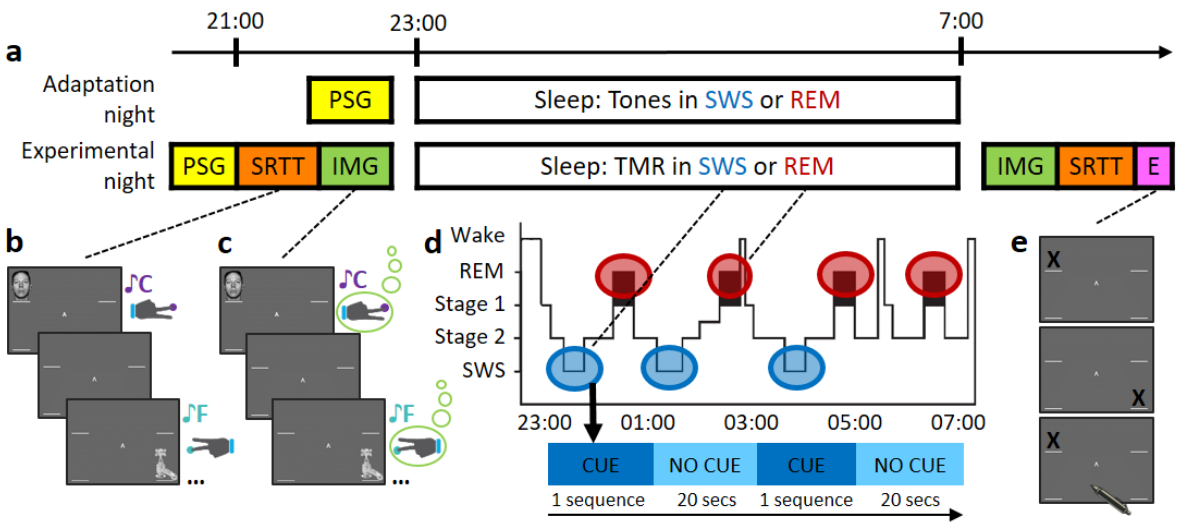


Figure 2.1. *Experimental procedures.* ***a)*** *The experiment consisted of an adaptation and an experimental night. During the adaptation night, participants were wired-up for EEG and while they slept tones were played as outlined in d). During the experimental night, participants were wired-up, after which they completed the SRTT and IMG tasks as outlined in b) and c), respectively. Then, participants went to sleep and TMR was carried out as described in d). After waking up, participants completed the IMG and SRTT again, and finally the explicit recall task (E) which is described in e).* ***b)*** *Serial Reaction Time Task (SRTT). Images were presented in two different sequences. Each image was accompanied by a specific pure tone (different for each sequence) and required a specific button press.* ***c)*** *Motor imagery task (IMG). Participants viewed and heard the same sequences again, but this time were instructed to only imagine pressing the buttons.* ***d)*** *Schematic representation of the TMR protocol. Reactivation took place in either SWS (blue bubbles) or REM sleep (red*

bubbles). One sequence was played as long as participants were in the relevant sleep stage, with a 20 second pause between repetitions. ***e)*** *Explicit recall task. Participants marked the order of each sequence on paper.*

Serial Reaction Time Task (SRTT)

The main task was a serial reaction time task (SRTT; adapted from Cousins et al., 2014; see Figure 2.1b) which contained two sequences of twelve items that were learned in interleaved blocks. The sequences – A: 1 2 1 4 2 3 4 1 3 2 4 3 and B: 2 4 3 2 3 1 4 2 3 1 4 1 – had been matched for learning difficulty, did not share strings of more than four items, and both contained each item three times. The blocks were interleaved so that a block of the same sequence was presented no more than twice in a row, and each block contained three repetitions of a sequence. There were 24 blocks of each sequence (48 blocks in total), and each block was followed by a pause of 15 seconds wherein feedback on reaction time (RT) and error-rate were presented. The pause could be extended by the participants if they wanted. Participants were aware that there were two twelve-item sequences and each sequence was indicated with 'A' or 'B' appearing centrally on the screen, but participants were not asked to learn the sequences explicitly. Counterbalancing across participants determined whether sequence A or B was the first block, and which of the sequences was cued during sleep.

Each sequence was paired with a group of pure musical tones, either low tones within the 4^{th} octave (C/D/E/F) or high tones within the 5^{th} octave (A/B/C#/D). These tone groups were counterbalanced across sequences. For each trial, a 200ms tone was played, and at the same time a visual cue appeared in one of the corners of the screen. The location indicated which key on the keyboard needed to be pressed as quickly and accurately as possible: 1 – top left corner = left shift; 2 – bottom left corner = left Ctrl; 3 – top right corner = up arrow; 4 – bottom right corner = down arrow. Participants were instructed to keep individual fingers of their left and right hand on the left and right response keys, respectively. Visual cues were neutral objects or faces, used in previous studies (Cousins et al., 2014, 2016). Stimuli appeared in the same position for each sequence (1 = male face, 2 = lamp, 3 = female face, 4 = water tap) and participants were instructed that the nature of the cues (objects/faces) was irrelevant. Visual cues stayed on the screen until the correct key was pressed, after which an 880ms inter-trial interval followed.

After the 48 blocks of sequences A and B, participants performed four more blocks that contained semi-random sequences which followed only the rule that no item was presented twice in a row. They contained the same visual stimuli and an 'R' displayed centrally on the screen. Two of these blocks were paired with the tone group of one sequence (cued in sleep), and the other two were paired with the tone group of the other sequence (not cued).

Motor Imagery Task (IMG)

After completion of the SRTT, participants were asked to do the same task again, but were instructed to only imagine pressing the buttons (Figure 2.1c). This task consisted of 30 interleaved blocks (15 of each sequence), presented in the same order as during the SRTT. Again, each trial consisted of a 200ms tone and a visual stimulus, the latter being shown for 880ms and followed by a 270ms inter-trial interval. There were no random blocks during this motor imagery task (IMG) and no performance feedback was presented during the pause between blocks.

TMR during REM and SWS

Cueing was started when participants – depending on their assigned group – were in stable REM sleep or SWS (fitting standard AASM criteria for Stage R or N3). Tones were presented as often as possible, with a pause of 1500ms between tones. One repetition of the sequence was presented, alternated by a 20 second break. Figure 2.1d shows a schematic representation of the TMR protocol. Cueing was paused immediately when participants showed any sign of arousal or when they left the relevant sleep stage. When a return to stable SWS/REM sleep was apparent, cueing was continued.

In the REM group, on average 984 sounds were played in the adaptation night (range 540-1404) and 1177 sounds in the experimental night (range 768-1718). In the SWS group, an average of 1159 sounds were played in the adaptation night (range 338-1836) and 1272 sounds in the experimental night (range 978-1908). There was no significant difference between the groups during the adaptation night ($t(30) = -1.37$; $p = 0.181$) or the experimental night ($t(30) = -0.97$; $p = 0.342$), as tested with two independent samples t-tests.

2.3.3 PSG data acquisition and analysis

Twenty-one electrodes were placed on the scalp and face of the participants following the 10-20 system. On the scalp, these were at 13 standard locations: Fz, Cz, Pz, F3, F4, C5, CP3, C6, CP4, P7, P8, O1, and O2, and they were referenced to the mean of the left and right mastoid electrodes. Further electrodes used were the left and right EOG, three EMG electrodes on the chin, and the ground electrode on the forehead. The impedance was <5kΩ for each scalp electrode, and <10kΩ for each face electrode. Recordings were made with an Embla N7000 amplifier and RemLogic 1.1 PSG Software (Natus Medical Incorporated). PSG recordings were manually scored by two trained

sleep scorers according to the standard AASM criteria (Berry et al., 2015). Both scorers were blind to the periods cueing occurred.

2.3.4 Electrophysiological analysis

The EEG data that were collected using the methods described above were further analysed with MATLAB (version R2016b) and the FieldTrip Toolbox (version 20/08/2019, Oostenveld, Fries, Maris, & Schoffelen, 2011). First, the raw data was coupled to the identity of the sounds played during sleep. Then, these marked continuous data were filtered between 0.1 and 30 Hz. Filter settings were based on previous studies (Schreiner et al., 2015; Schreiner & Rasch, 2015). Data were then coupled to the sleep scoring. This allowed for the removal of any trials that were played during the wrong sleep stage or during an arousal, since arousals had been marked visually during sleep scoring. Subsequently, the continuous data were segmented into trials starting one second before sound onset and ending three seconds after sound onset. Trials that took place during the wrong stage or during an arousal were discarded. In the adaptation night, this resulted in the removal of on average 5.8% and 6.4% of data in the REM and SWS groups, respectively. In the experimental night, this resulted in the removal of on average 4% of data in the REM and 10.7% of data in the SWS group.

Further artifacts were removed in a multi-step procedure. These steps were based on recommendations from the literature (Cohen, 2014; Oostenveld et al., 2011) and informed by the particular challenges of the data (e.g. the use of ICA in the REM data). Trials were first re-segmented into smaller trials of -0.5 and +3 seconds around the onset of a sound. Each EEG channel was then looked at separately. A trial was considered an outlier for a given channel if it was more than two standard deviations from the mean on amplitude or variance. If a trial was considered an outlier in more than 25% of channels (i.e. 3 channels), then this trial was rejected. During this procedure, on average 10.5% of trials were removed from the adaptation night data in the REM group, and 11.9% of trials in the SWS group. In the experimental night, this led to removal of 9.8% of trials in the REM, and 11.7% of trials in the SWS group. Those trials which were considered an outlier in <25% of channels were interpolated based on triangulation of neighbouring channels. Data in the REM group was subsequently analysed with independent component analysis (ICA), to remove eye movement artifacts which can occur during REM. Components identified by the ICA were correlated with the signal from the eye electrodes, and components that were significantly correlated (corrected for multiple comparisons) were removed. In the final artifact rejection step, all channels for each participant were manually inspected. This is common in the literature (e.g. Cairney et al.,

2018; Göldi, van Poppel, Rasch, & Schreiner, 2019; Lehmann et al., 2016; Schreiner et al., 2015). Any channels that showed overall noise were interpolated based on their neighbours, and overly noisy trials still detected during this step were removed. After artifact rejection, approximately 766 trials were left for analyses in the adaptation night (range 371-1188) and 938 trials in the experimental night (range 502-1418) in the REM group. In the SWS group, there were 954 trials left in the adaptation night (range 272-1481) and 1018 trials in the experimental night (range 646-1381).

Event-related potentials (ERPs) were analysed time-locked to TMR cue start. To reduce influence of outliers, we used the median to calculate the ERP of the segments. Grand-averages were baseline-corrected to a baseline window of -1 second until cue onset. This long baseline window was chosen because of the low-frequency nature of slow-wave sleep and based on the literature (Schreiner et al., 2015). ERPs were expected to occur shortly after cue onset, and therefore statistical analyses focused on the time period between cue onset and 500ms thereafter. Due to the pronounced difference in electrophysiology between SWS and REM, comparisons between cues (left- versus right-handed cues, and cues during the adaptation versus experimental night) were done separately for the SWS and REM groups. Like in previous research (Cairney et al., 2018), they were performed as paired-samples t-tests and corrected for multiple comparisons using FieldTrip's nonparametric cluster-based permutation method, using 1000 permutations. Results were considered significant at $p < 0.05$.

2.3.5 Spindle analysis

Previous studies using similar tasks have indicated a relationship between sleep spindles (short bursts of activity in the 11-16Hz range) over central or frontal electrodes and behavioural outcomes (Antony et al., 2012; Cousins et al., 2014). Therefore, spindles in the SWS group were also examined here, using a counting algorithm based on one used by Antony and colleagues (Antony et al., 2018). In short, the raw EEG was filtered in the sigma band as identified in the paper by Antony et al. (11-16Hz), and root-mean-square (RMS) values were calculated using a sliding 200ms window. Any segments that fit threshold criteria were selected. To fit threshold, a spindle had to be during stage N3 (SWS), be between 0.3 and 2.5 seconds long, and have at least 5 oscillations in that period. Stage N3 was selected because this was the stage during which our cues were presented, allowing us to investigate spindles related to TMR cues specifically. Length and oscillation criteria were based on a previous study (Navarrete, Schneider, et al., 2020). Spindle identity was then further confirmed

by using time-frequency information to detect whether increased spindle-band power indeed took place during the selected segments (Navarrete, Schneider, et al., 2020; Purcell et al., 2017).

After detection, any spindles that fell partly or wholly during a previously visually marked arousal were removed. The remaining spindles were divided into those taking place during TMR, and those taking place during baseline SWS. If a spindle started up to 1.65 seconds after tone onset, it was considered to occur during TMR. This interval was chosen because subsequent tones in the sequence, when a sequence was played without pauses, would be a maximum of 1.65 seconds apart. If the start of a spindle fell outside of this interval, it was considered to occur during baseline. Based on the analyses and results of Cousins and colleagues (2014) we determined spindle density and laterality. Spindle density was calculated as the amount of 'within TMR' spindles divided by the amount of cues played during the night. We then computed spindle laterality in the left hemisphere by subtracting spindle density in the right from the left hemisphere. To obtain spindle laterality in the right hemisphere, finally, we subtracted spindle density in the left from that in the right hemisphere. Because Cousins and colleagues (2014) found an effect of spindle laterality in central electrodes on procedural skill improvement in the SRTT, our analyses focused on electrodes C5 & CP3 (left hemisphere) and C6 & CP4 (right hemisphere). Notably, because our participants used both hands, while the participants in prior studies showing a relationship between spindle laterality and finger tapping improvement across sleep have used just one hand (Antony et al., 2012; Cousins et al., 2014, 2016; Walker, Brakefield, et al., 2002), we did not necessarily expect to find a relationship.

2.3.6 Behavioural analysis

Performance on the SRTT was measured by the reaction time (RT) per block. Following the method used by Cousins et al. (Cousins et al., 2014, 2016), any trials with an RT of more than 1000ms were excluded from analyses, while trials with incorrect button presses prior to the correct ones were not excluded. Because we were interested in differences between the dominant and non-dominant hands, as well as overall performance, the data were analysed in three different ways: Both Hands (BH), Left Hand (LH), and Right Hand (RH).

Subjects were excluded if (1) their RT performance before sleep was >2 SDs from the group mean (n = 1 in the LH dataset), (2) there was a >2 SD disparity between the group mean RT for the two sequences before sleep (n = 1 in the BH dataset, n = 2 in the LH dataset, and n = 2 in the RH dataset), or (3) they exhibited a positive slope of the learning curve before sleep, i.e. they did not show any

learning during training (*n* = 1 for all datasets). Thus, thirty participants remained in the BH dataset (*n* = 15 for the REM group, and *n* = 15 for the SWS group), twenty-eight in the LH dataset (*n* = 15 for the REM group, and *n* = 13 for the SWS group), and twenty-nine in the RH dataset (*n* = 15 for the REM group, and *n* = 14 for the SWS group).

RTs were divided into those for the cued and those for the uncued sequence. Performance on the last four blocks before sleep was considered to represent pre-sleep ability, and this was subtracted from the random blocks to remove the effects of increased sensorimotor mapping ability. The resulting variable can thus be called sequence-specific skill. *Sequence specific improvement* was then calculated for each sequence by subtracting the pre-sleep sequence-specific skill from the post-sleep sequence specific skill (i.e. the random blocks after sleep minus the first 4 blocks after sleep). Higher values on this value thus indicate more improvement.

We used analyses of variance (ANOVAs) to determine the effects of time (before/after sleep) and sequence (cued/uncued) within each group (REM/SWS) separately. Note that the Shapiro-Wilk test revealed statistically significant deviations from normality in some of the sub-groups, though plotting showed that these deviations were small. Straightforward non-parametric alternatives to ANOVA (e.g. Friedman's ANOVA) do not allow for testing of interaction effects, like an interaction of time and sequence which would indicate an effect of TMR. With this in mind, and given the fact that the deviations from normality were small, we chose to use and report parametric ANOVAs for the initial evaluation of the data. When follow-ups of interactions were required, we used both paired-sample t-tests and Wilcoxon signed-rank tests to determine the reliability of the results. These tests always led to the same conclusions in terms of significance or non-significance of observed differences – and therefore we report only one of them in the results section. Specifically, we will report the Wilcoxon signed-rank tests, because this is the more conservative test. Relationships between behavioural measures and features of the EEG and sleep were assessed with Pearson's correlations, or Spearman's Rho in the case of non-normal distributions. All statistical tests were 2-tailed and considered significant for $p < 0.05$. Analyses were conducted in R (version 3.6.3, R Core Team, 2020). We included measures of effect size: generalised eta squared (η^2_G) for ANOVA as calculated with the "afex" R package (Bakeman, 2005; Lakens, 2013; Olejnik & Algina, 2003; Singmann, Bolker, Westfall, Aust, & Ben-Shachar, 2020), and *r* for Wilcoxon tests as calculated with the "rcompanion" R package (Fritz, Morris, & Richler, 2012; Mangiafico, 2020).

2.3.7 Classification

We trained an EEG classifier to classify right- versus left-handed trials. This classifier was trained and tested using data from the motor imagery task (IMG) performed just before and after sleep to assess the classifier performance during wake. We started by band pass filtering the EEG signal from 0.1 to 50 Hz, and averaging 80ms time bins (40ms before and 40ms after each individual time point). This 80ms smoothing window was applied to the whole trial. The resulting ERP features were submitted to a Linear Discriminant Analysis (LDA) classifier. A time x time classification was then performed using features from one time point to train a classifier, and that classifier was then tested on all time points. These steps were based on recommendations for time x time classification in the literature (Dehaene & King, 2016). Because wake-to-sleep classification of EEG data is in its infancy, the remaining steps were developed by reasoning about and devising logical solutions to the specific challenges present in sleep data in general, and our data in particular.

We reasoned that if the classifier did not perform well in wake (either because the memory is weakly encoded or because it can somehow not classify the encoded memory), then it would not work during sleep, where noise is much higher and signal much lower. We therefore used classifier performance during wake as a filter, and excluded participants in whom wakeful reactivation could not be classified above 0.7 correct rate from further classification. During memory reactivation after a cue, there may be a time where activation reaches a peak, and other time points may not be very relevant for classification. We therefore used wakeful classification to extract the time period when classification accuracy was highest. This 'peak activation period' is very important for classification. We defined this time period as the time of interest (TOI). Using our wake-to-wake classifier, we identified a TOI based on the time of the highest classification rates. This is the window when we can best discriminate between the two classes, defined using a threshold of 0.75 correct on the grand average accuracies of all participants.

Subsequently, we developed an EEG classifier using wake samples and applied it on sleep. This was trained using every time point of wake and applied on sleep after each TMR cue. If reactivation occurs during sleep then we would expect the classification to peak when we train the classifier with the time points of the TOI that we identified during wake motor imagery. We applied the classifier to data from both adaptation and experimental nights for REM and SWS groups, as the comparison between these two nights allows us to separate the brain response to sounds (adaptation night) from the brain response to memory-related cues (experimental night). If TMR is associated with genuine memory reactivation, classification should be stronger during the

experimental night, when participants have associated tones with the task, than during the adaptation night when tones have no memory associations.

We devised a method for removing noisy trials and keeping the good ones. In this method, trials which had low posterior probability (i.e. those which fell near the decision boundary) were considered noise and eliminated from the analysis. Rather than defining a set cut-off value, we used the maximum number of trials that was available for all participants consistently to determine which trials would be kept. In the SWS group this meant 300 trials per participant, and 366 in the REM group. In other words, we used the trials which the classifier was most confident about to assess classifier performance. Importantly, this process does not consider the actual class label – it only considers the distance from the decision boundary. The exact same process was employed for classification of both the experimental and adaptation nights. After we had removed these noisy trials, classification accuracy on experimental and adaptation nights was compared to determine whether the classifier was detecting memory reactivation.

Given that the SRTT is a motor task and we are classifying right and left hand presses, we expected to obtain more meaningful results by focusing on the motor area when obtaining features. Thus, we repeated the classification analysis using only the four channels around the motor area: CP3, C5, CP4, and C6. This final classification pipeline is shown in Figure 2.2. It uses the TOI as identified with the classification using all channels. However, whereas previously each time point had a classification output, here we aggregated the time points inside the TOI together on motor channels to form feature vectors. This allows the classifier to take into account more information, which should enable it to learn the classes better. Put differently, this analysis only provides one overall classification for a trial, rather than one for each time point, but this classification has higher confidence because it is based on more information. Signals were band pass filtered from 0.1 to 50 Hz, and ERP features were extracted and aggregated from the sleep TOI and the chosen channels and then fed to the LDA classifier.

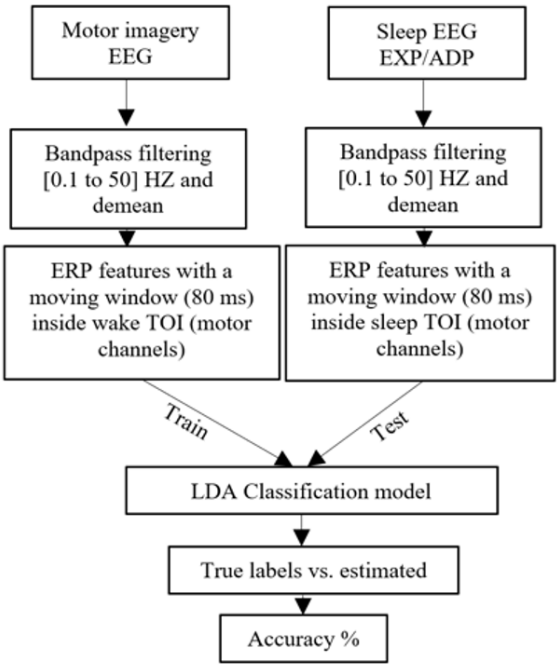


Figure 2.2. *Block diagram of the final classification pipeline. Signals were band pass filtered, and smoothed time points inside the TOI were then aggregated using motor channels to form feature vectors that were subsequently given to the classifier for classification.*

2.4 Results

2.4.1 Sleep parameters

Sleep scoring confirmed that the vast majority of all TMR sounds were played during the correct sleep stage. No sounds were played in the opposite sleep stage (i.e. during SWS for the REM group, and vice versa). Sleep parameters did not differ between groups, with the notable exception of Stage 2, which was longer in the SWS group during both the adaptation night ($F(1,30) = 6.2$; $p = 0.018$) and the experimental night ($F(1,30) = 6.1$; $p = 0.020$). A summary of the time spent in sleep stages can be found in Table 2.1.

We combined the groups and conducted paired t-tests to see whether sleep was better on the second night spent in the lab. Participants slept significantly longer in the experimental compared to the adaptation night ($t(31) = -2.23$; $p = 0.033$, see Table 2.1). Participants also spent more time in REM in the experimental night than the adaptation night ($t(31) = -2.16$; $p = 0.038$), but the difference in Stage 2 sleep did not reach significance ($t(31) = -1.84$; $p = 0.075$). There was also no difference in SWS ($t(31) = 0.41$; $p = 0.679$) or Stage 1 sleep ($t(31) = 1.66$; $p = 0.107$). On the other hand, participants did wake up more during the adaptation night compared to the experimental

night (t(31) = 2.42; p = 0.022), which may contribute to the difference in total sleep time across the nights.

Table 2.1. *Average minutes spent in sleep stages (± standard deviation). * = significant at the α = 0.05 level.*

	REM sleep group *n = 16*		**SWS group** *n = 16*		**Significance values**
	Adaptation night	**Experimental night**	**Adaptation night**	**Experimental night**	**ADP versus EXP night**
Stage 1	30.2 ± 15.6	27.8 ± 13.3	32.6 ± 24.3	26.8 ± 19.2	p = 0.107
Stage 2	197.3 ± 29.8	207.4 ± 22.2	223.9 ± 30.5	234.2 ± 37.4	p = 0.075
SWS	119.8 ± 34.2	116.8 ± 21.5	109.9 ± 33.5	109.9 ± 33.1	p = 0.679
REM	96.6 ± 24.5	103.4 ± 24.5	92.6 ± 31.8	104.5 ± 24.8	p = 0.038*
Total sleep time	444.0 ± 32.9	455.4 ± 34.1	458.9 ± 34.8	475.4 ± 29.9	p = 0.033*
Wake after sleep onset	25.0 ± 24.5	16.2 ± 15.8	23.5 ± 18.5	14.4 ± 15.4	p = 0.022*

We also evaluated self-rated quality of sleep across nights, and self-rated positive feeling on the mornings after waking up in the lab. In terms of sleep quality, on average this was rated higher after the experimental compared to the adaptation night (t(31) = -2.32; p = 0.027). However, when evaluating the ratings of how participants felt in the morning, we found no difference between the nights (t(31) = 0.25; p = 0.808).

2.4.2 Behaviour: Both Hands

We first evaluated behavioural results for both hands combined. At the end of pre-sleep training, RTs were significantly faster for trials within a sequence compared to trials within the random blocks, confirming learning of both sequences in both the REM and the SWS groups, (all p < 0.001). Importantly, during pre-sleep baseline, RTs did not differ between the cued and uncued sequences for either sequence or random blocks (lowest p: F(1,28) = 1.22; p = 0.280).

An ANOVA with factors Time (before and after sleep) and TMR (cued or uncued sequence) examined sequence-specific skill change in SWS and REM groups. In SWS, as expected, there was a main effect of time (F(1,14) = 11.35; p = 0.005; η_G^2 = 0.108), with faster performance after sleep. There was also a time x TMR interaction (F(1,14) = 5.01; p = 0.042; η_G^2 = 0.015), with more improvement shown on the cued versus the uncued sequence, as determined with a Wilcoxon

signed-rank test (V = 84; p = 0.049; r = 0.529; pictured in Figure 2.3e). The main effect of TMR across pre and post-sleep sessions was not significant (F(1,14) = 0.33; p = 0.574). In the REM group, our ANOVA showed a main effect of time (F(1,14) = 44.77; p < 0.001; η_G^2 = 0.137), but no other main effects or interactions (lowest p: F(1,14) = 0.37; p = 0.553), Figure 2.3e.

2.4.3 Behaviour: Left and Right Hand

To investigate whether TMR influenced each hand similarly, we separated trials into those where responses were made with the left (non-dominant) or right (dominant) hand. At the end of pre-sleep training performance of both cued and uncued sequences was significantly faster than performance on the random blocks (all p < 0.001) in both right hand (RH) and left hand (LH). This confirms learning in both the LH and RH before sleep. Before sleep RTs in either hand did not differ between the cued and uncued sequences, nor between random blocks before sleep (lowest p: F(1,27) = 1.93; p = 0.176), see Figure 2.3a-d.

We analysed sequence-specific skill change in the REM and SWS groups separately for RH and LH using an ANOVA with within-participant factors time (before and after sleep) and TMR (cued or uncued sequence). Starting with the LH, in the SWS group, this showed a main effect of time (F(1,12) = 9.09; p = 0.011; η_G^2 = 0.092) and a time x TMR interaction (F(1,12) = 7.31; p = 0.019; η_G^2 = 0.015). A Wilcoxon signed-rank test (V = 79; p = 0.017; r = 0.649) showed greater overnight improvement on the cued than the uncued sequence, Figure 2.3f. The main effect of TMR across both sessions was not significant (F(1,12) = 1.29; p = 0.279; η_G^2 = 0.011). In the REM group, LH showed a main effect of time (F(1,14) = 25.36; p < 0.001; η_G^2 = 0.110), but no other effects (lowest p: F(1,14) = 0.80; p = 0.387).

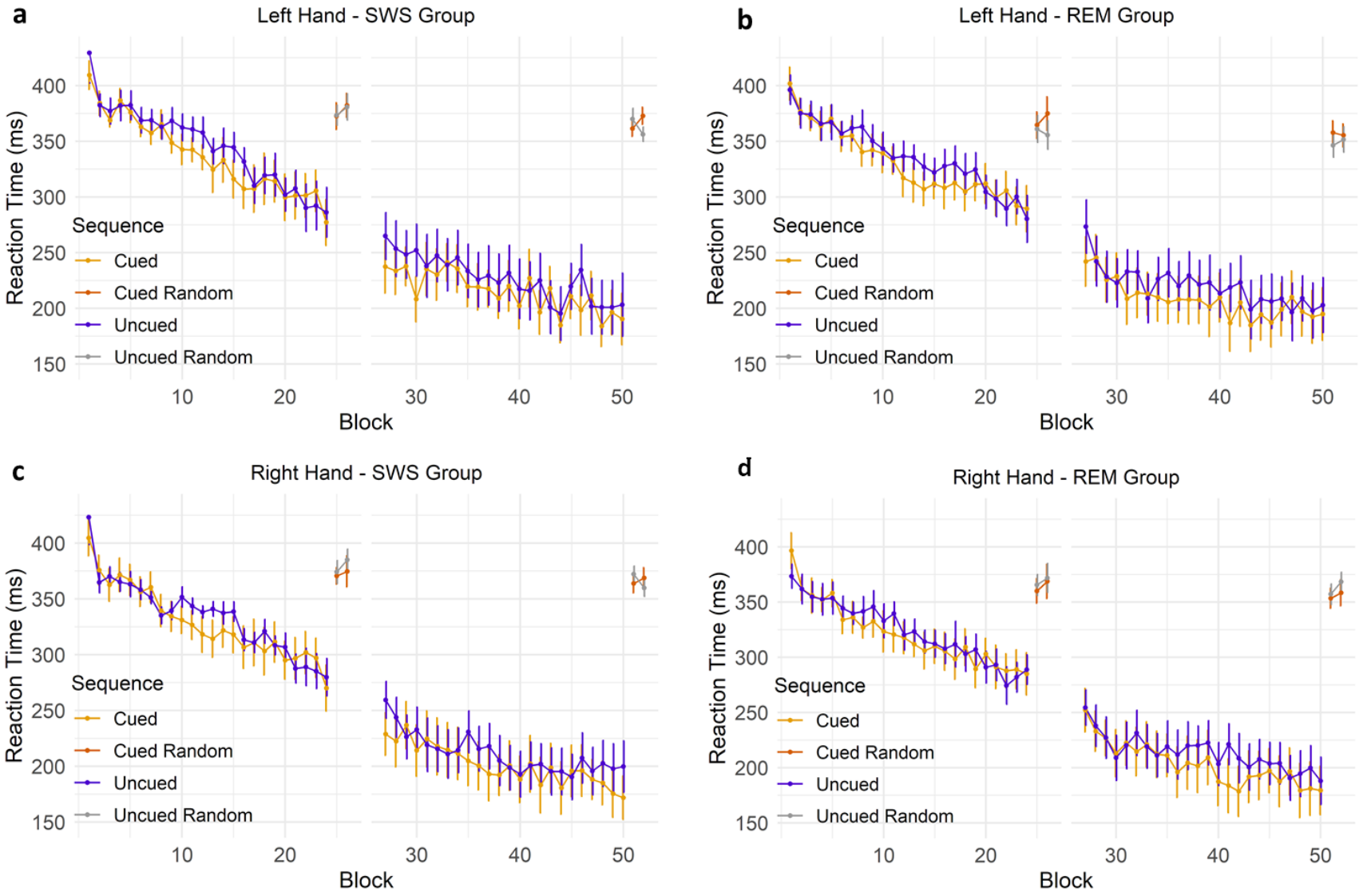
Turning to the RH, this showed no cueing-related improvement in either SWS or REM groups. In the SWS group the main effect of time was significant (F(1,13) = 8.63; p = 0.012; η_G^2 = 0.113), but there were no other main effects or interactions (lowest p: F(1,13) = 2.98, p = 0.108; η_G^2 = 0.012; for the time x TMR interaction). In REM, there was a main effect of time (F(1,14) = 44.47; p < 0.001; η_G^2 = 0.144), but no other main effects or interactions (lowest p: F(1,14) = 1.77, p = 0.205). Figure 2.3g shows the sequence-specific improvement for the cued and uncued sequences per group in the RH.

Given the differences in TMR efficacy in right and left hand, we were interested to know whether either hand showed a weaker performance from the outset. We conducted a mixed ANOVA with within-participant factor hand (left or right) and between-participant factor group (REM or SWS) on

sequence-specific skill in the last four blocks before sleep. There was a trend towards an effect of hand ($F(1,24) = 3.88$, $p = 0.060$; $\eta_G^2 = 0.011$), with the left hand showing weaker sequence-specific skill than the right hand. There were no other main effects or interactions (lowest p: $F(1,24) = 0.35$, $p = 0.558$). Because effects related to handedness might be clearer in the reaction time compared to the sequence-specific skill, we also analysed this raw RT. Using the same mixed ANOVA on raw RT in the last four blocks before sleep, there was a main effect of hand ($F(1,24) = 15.59$, $p < 0.001$; $\eta_G^2 = 0.024$). There were no other main effects or interactions (lowest p: $F(1,24) = 0.63$, $p = 0.437$). Overall, the left hand thus showed lower sequence-specific skill and slower raw RT.

2.4.4 Electrophysiology: Spindle analysis

To determine whether cued sequence advantage related to spindles, we calculated a 'procedural cueing benefit' by subtracting cued from uncued sequence reaction time in the first four blocks after sleep (Cousins et al., 2014). Thus, higher values denote a faster RT in the cued compared to the uncued sequence (i.e. a cueing benefit). However, there were no significant correlations between spindles and cueing benefit.


a
Left Hand - SWS Group
b
Left Hand - REM Group
c
Right Hand - SWS Group
d
Right Hand - REM Group
Reaction Time (ms)
Block
Sequence
Cued
Cued Random
Uncued
Uncued Random
150
200
250
300
350
400
10
20
30
40
50

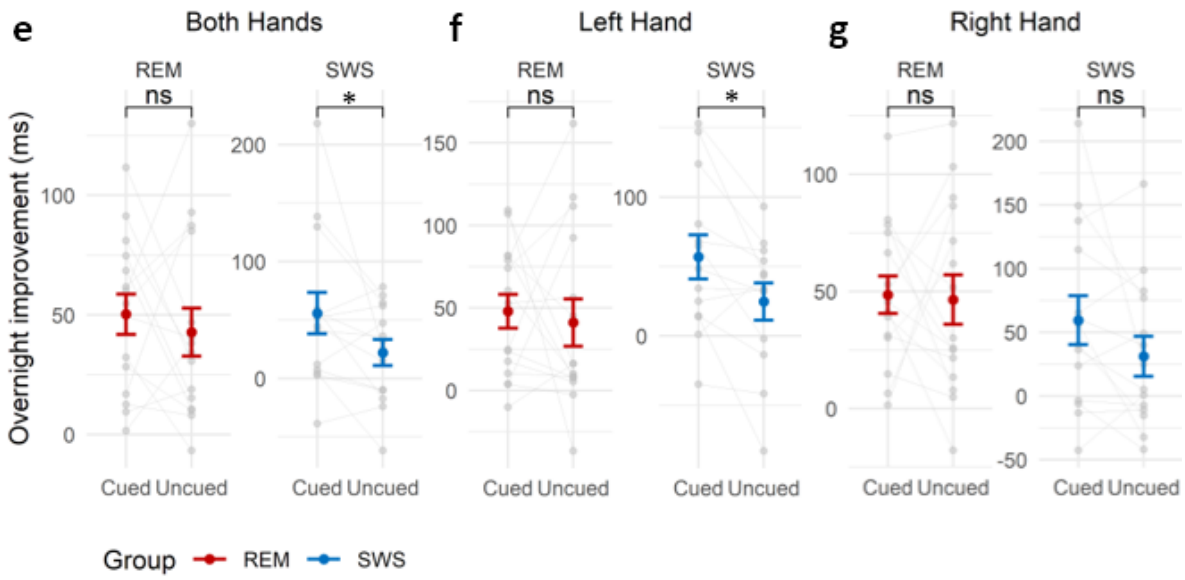


Figure 2.3. *Behavioural results. * = p < 0.05.* ***a)*** *SRTT performance across all blocks of learning, for the SWS group in the left hand.* ***b)*** *SRTT performance across all blocks of learning, for the REM group in the left hand.* ***c)*** *SRTT performance across all blocks of learning, for the SWS group in the right hand.* ***d)*** *SRTT performance across all blocks of learning, for the REM group in the right hand.* ***e)*** *SRTT overnight sequence-specific improvement in both hands was significantly better for the cued than uncued sequence in the SWS group only (V = 84; p = 0.049).* ***f)*** *SRTT overnight sequence-specific improvement in the left hand was greater for the cued than uncued sequence in the SWS group only (V = 79; p = 0.017).* ***g)*** *SRTT overnight sequence-specific improvement in the right hand did not differ between cued and uncued sequence in either group (the difference in the SWS group is not significant at V = 77; p = 0.135). Data are presented as mean ± SEM.*

2.4.5 Electrophysiology: Event-related potentials

We examined event-related potentials (ERPs) in response to left- and right-handed TMR cues in sleep for the SWS and REM groups. Interestingly, TMR in REM elicited a stronger response to right-hand compared to left-hand cues when all EEG channels were combined in the experimental night. A cluster-based permutation test on the combined EEG channels and the latency range of 0 to 500 ms post-stimulus showed that this difference was significant between 0.225 and 0.285 seconds after cue onset (p = 0.044; see Figure 2.4a). Though this difference in right and left hand cues was visible throughout the brain, it was descriptively most pronounced in the left hemisphere (C5 and CP3). In the adaptation night there was no difference between cues associated with the left versus the right hand (lowest p = 0.238). Although TMR in SWS revealed a numerical trend in keeping with the REM results, there were no significant differences in responses to right and left hands in either experimental (lowest p = 0.447) or adaptation, (lowest p = 0.311) nights. Descriptively, it is interesting to note that TMR cues in this sleep stage showed the slow-oscillatory pattern that is

characteristic of slow-wave sleep (see Figure 2.4b). We further examined ERPs in each hemisphere separately, but this did not reveal any additional information. Supplementary analyses comparing ERPs on the adaptation and experimental night, as well as time-frequency analyses, can be found on page 75.

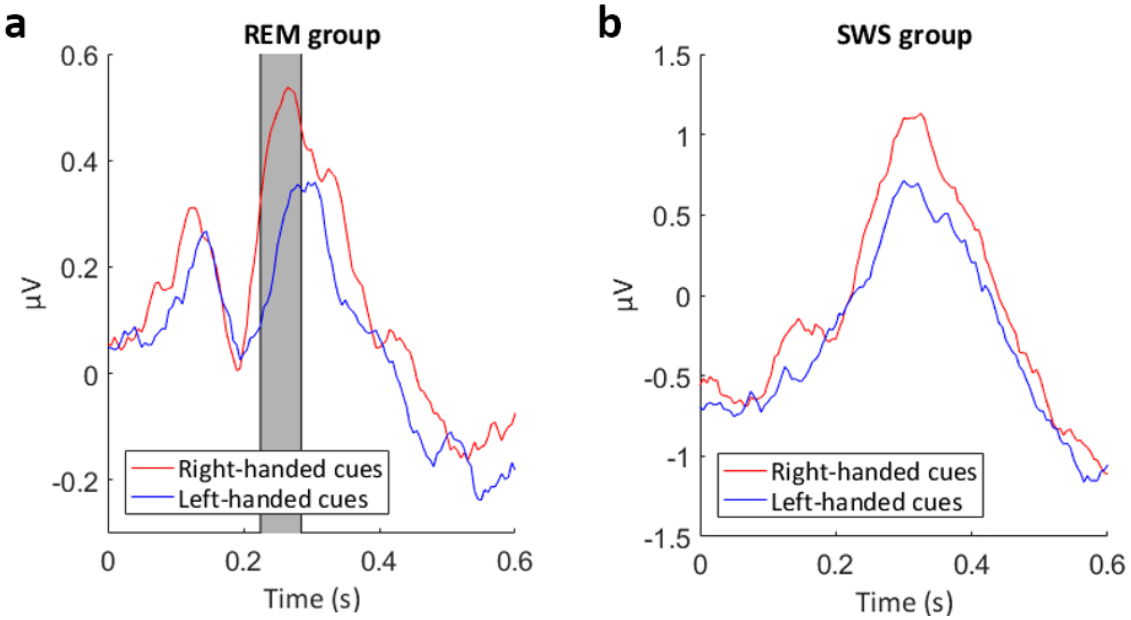


Figure 2.4. *ERPs to left- and right-handed cues over all EEG channels during the experimental night* ***a)*** *In the REM group. Grey rectangle indicates when the difference between these cues is significant ($p < 0.05$; corrected for multiple comparisons).* ***b)*** *In the SWS group.*

2.4.6 Classification

We first produced a time x time classification in wake, Figure 2.5. Here, the accuracy for classifying left- versus right-handed trials using a classifier trained at the specified 'training time' and tested at the specified 'testing time' was built up one row at a time. Using the correct rate threshold of 0.75 that we defined (see Classification methods), this revealed a time of interest (TOI) from 0.7 to 1.1 seconds after cue onset for the SWS group (Figure 2.5a) and 0.64 to 0.97 seconds after cue onset for the REM group (Figure 2.5b).

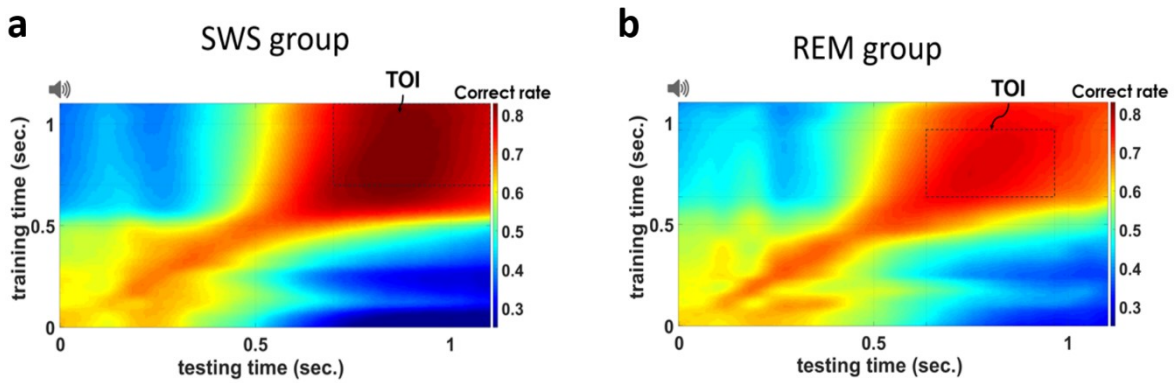


***Figure 2.5**. Classification in wake: Grand average classification accuracy of wake motor imagery of right hand vs. left hand (wake imagery training and testing) using ERP features with an 80ms smoothing window and LDA classifier. For **a)** SWS group and **b)** REM group.*

Turning to the sleep data, we next examined classification in sleep with a second time x time classification procedure, but this time training with wake (y-axis) and testing with sleep (x-axis), Figure 2.6. To avoid double-dipping, we used a leave-one-subject-out approach wherein, for each subject, data from the other subjects was used to select a sleep TOI for that subject. An example of the sleep classification result and the calculation of the sleep TOI for one participant from the SWS group is shown in Figure 2.6. Note that the TOI in sleep was defined as the window with the highest accuracy, where window length was determined by the window length obtained during wake classification. The TOI during sleep varied slightly between participants but interestingly, SWS TOI occurred later for the experimental night than it did during wake, from 0.88 ± 0.04 to 1.28 ± 0.04 seconds after cue onset. The exact same analysis was performed on the adaptation night to get its TOI of 0.55 ± 0.036 to 0.95 ± 0.036 seconds after cue onset. For the REM group the TOI for the experimental night was 0.74 ± 0.004 to 1.07 ± 0.004 seconds after cue onset, and the TOI of the adaptation night was 1.1 ± 0.21 to 1.4 ± 0.21 seconds after cue onset.

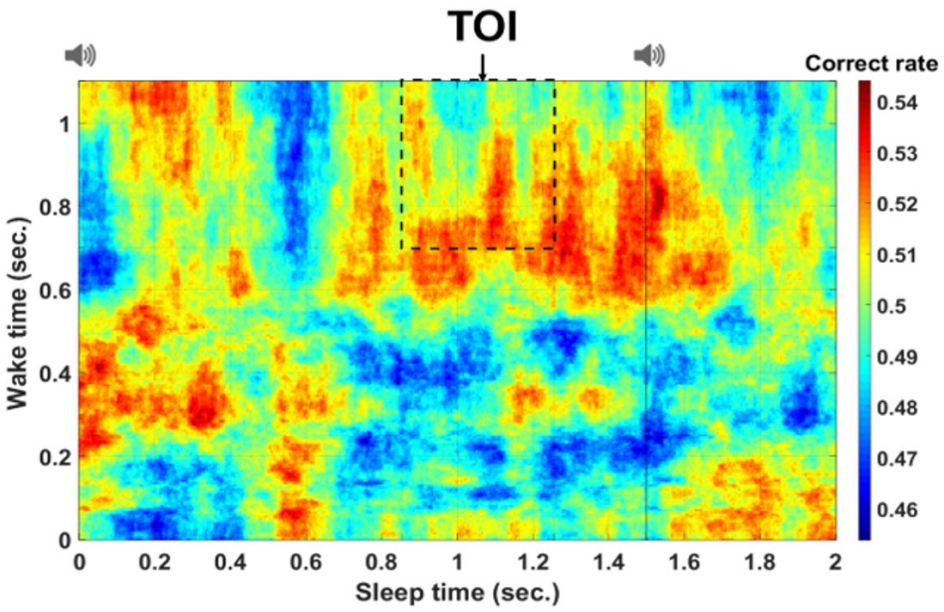


Figure 2.6. *Classification of right- versus left-handed trials during the experimental (EXP) night. Note that the classifier was trained during wake (time of training shown on the Y axis) and applied during sleep (time of application shown on the X axis). The TOI for one participant from the SWS group is shown here. TOI is marked with a dashed square and is calculated by leaving the data of that participant out and getting the maximum window from average classification of all other participants of the same group.*

We applied our classifier to SWS and assessed classification accuracies using 300 trials, as this was the maximum number we could use consistently across all participants. Figure 2.7a shows the accuracies for the experimental versus the adaptation data for the SWS group. As mentioned in the methods section, we only applied our classifier to sleep in those participants that classified above 70% in wake. This was done because we reasoned that if the classifier did not perform well in wake (either because the memory is weakly encoded or because it can somehow not classify the encoded memory), then it would not work during sleep, where noise is much higher and signal much lower. In SWS, classification accuracy was significantly higher for the experimental night than the adaptation night (paired t-test, $t(9) = 4.1$; $p = 0.003$). Classification also performed significantly above the chance level of 50% ($t(9) = 3.94$; $p = 0.003$). For completeness, we also extracted the TOI based on the adaptation night, which should be a time period that does not relate to the encoded memory of the hands, because the task had not yet been completed in this night. Thus, we would expect a non-significant difference between the nights if we use the TOI defined using the adaptation night. This was the case: a paired t-test showed no difference between the nights in this analysis, (paired t-test, $t(9) = -1.6$; $p = 0.14$). This suggests that only the experimental night contained memory-related reactivation that is similar to the encoded memory.

We repeated the process for the REM group for both experimental and adaptation nights. For this group we included 366 trials, as the maximum available for all participants. The classification performance did not exceed chance level and showed no difference between experimental and adaptation nights (paired t-test, t(13) = 1.57; p = 0.14) as shown in Figure 2.7b. Interestingly, however, if the outlier who obtained only 40% correct classification in the experimental night is rejected, the difference between experimental and adaptation nights becomes significant: mean accuracy for the experimental night: 51.6%, and for the adaptation night: 49.4%, (paired t-test, t(12) = 2.93; p = 0.013). Furthermore, the experimental night is significantly higher than chance level 50% (t(12) = 2.93; p = 0.013). As expected the classification using the TOI defined using the adaptation night did not show a significant difference between the nights (paired t-test, t(13) = 1.01; p = 0.332).

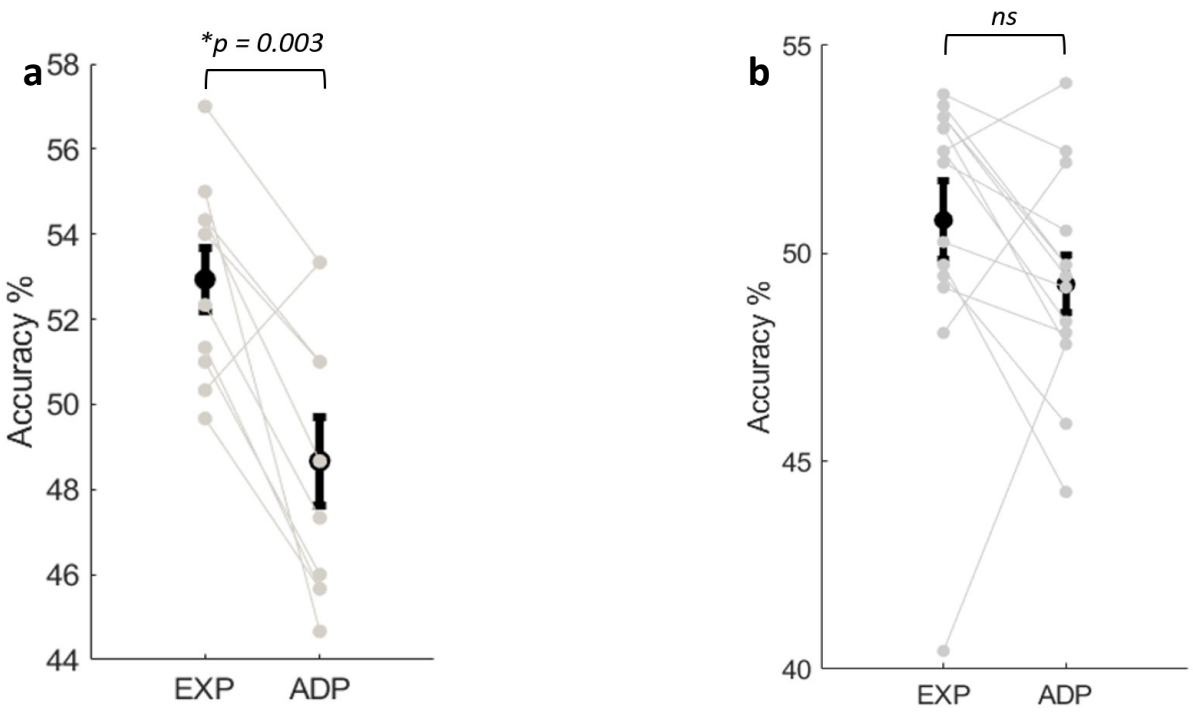


Figure 2.7. *Classification accuracy in the experimental and control (adaptation) night, for both the* ***a)*** *SWS group, (paired t-test, t(9) = 4.1; p = 0.003) and* ***b)*** *REM group, (paired t-test, t(13) = 1.57; p = 0.14), this difference becomes significant if the outlier point at ~40% is rejected (paired t-test, t(12) = 2.93; p = 0.013)*

2.4.7 Correlating classification performance with behaviour

We tested for correlations between the classification performance and behaviour improvement as summarised in Table 2.2, but there was no significant relationship with either early (first four blocks after sleep) or late (last four blocks after sleep) behavioural improvements for any of the groups.

Table 2.2. *Correlations between classification performance and either early or late behaviour improvement for both groups.*

	REM sleep group ***n = 14***		**SWS group** ***n = 10***	
Early improvement	r = -0.15	p = 0.63	r = -0.16	p = 0.66
Late improvement	r = 0.18	p = 0.55	r = 0.24	p = 0.51

2.5 Discussion

In this study, we set out to determine the effects of TMR during SWS and REM sleep on an SRTT. We additionally wanted to know whether this technique influences dominant and non-dominant hands equally. Finally, we also developed a machine learning classifier to investigate whether memory reactivation could be detected after TMR in SWS and REM sleep.

Our results demonstrate that auditory TMR in SWS benefits next-day performance on the SRTT. This is in line with previous studies (Cousins et al., 2014, 2016) and our hypothesis, which stated that we expected to replicate the finding that TMR during SWS would improve behaviour on the cued sequence as compared to the uncued sequence. In contrast, TMR during REM sleep did not improve behaviour on the cued compared to the uncued sequence, contrary to our hypothesis. When separating the results by hand, it became clear that TMR during SWS only benefitted cued sequence performance in the non-dominant hand. There was no significant difference between performance on the cued and uncued sequences in the dominant hand. This is in accordance with our hypothesis that the non-dominant hand would particularly benefit from TMR. However, TMR during REM sleep did not differentially affect the dominant and non-dominant hands, contrary to our expectations.

Looking at our classifier, we found that we were able to classify TMR during SWS. This classification was both significantly better than chance and significantly better in the experimental than the adaptation night. In other words, our expectation that we would be able to classify TMR during SWS was correct. Classification of TMR during REM was not significantly better than chance and not significantly better in the experimental than the adaptation night, except if a large outlier was removed. Note, however, that this was intended as an exploratory analysis to investigate the feasibility of our current method of classification during REM sleep.

ERP results indicated that the brain was able to distinguish between cues relating to the left and the right hand during REM sleep, as evidenced by a significant difference between these cues in the REM group. This is in line with our hypothesis that we would find differences between cues relating to the dominant and non-dominant hand in our electrophysiological results. However, these differences did not reach significance in the ERP results of the SWS group, nor in the time-frequency results of both of the groups. This is not consistent with our hypotheses. Finally, we found that benefits of TMR during SWS were not related to spindle laterality, contrary to previous findings (Cousins et al., 2014) and our hypothesis.

2.5.1 TMR in SWS but not REM benefits SRTT consolidation

Early studies of memory consolidation in sleep suggested that REM is critical for motor skills (Karni et al., 1994; Smith, 1993, 1995, 2001; Smith & Smith, 2003). Furthermore, an influential study showed reactivation of motor networks during REM after a finger tapping task (Maquet et al., 2000). These findings were subsequently extended by showing that learned material content and acquisition level before sleep modulated this REM reactivation (Peigneux et al., 2003). While TMR is not often applied in REM, a number of the studies that do exist have found significant effects, for instance on the affective tone of memories (Rihm & Rasch, 2015), on fear conditioning (Hars et al., 1985), on complex a logic task (Smith & Weeden, 1990), and on Morse code learning (Guerrien et al., 1989). Two studies of odour based TMR compared the impacts of TMR in REM and Stage 2 or SWS on a declarative procedural memory task (Laventure et al., 2016; Rasch et al., 2007). While both of these found an impact of TMR in NREM, neither showed an impact of REM TMR. Our findings are in keeping with these odour-based studies, since we found neither behavioural benefits of REM TMR nor reliably classifiable EEG responses. The fact that our EEG classifier detected reactivation in SWS but not REM builds on this to suggest that this task does not reactivate in REM sleep, or at least not in response to TMR cues. However, our own prior observation that the time spent in REM modulates neuroplasticity in the motor system as a result of TMR in SWS (Cousins et al., 2016) suggests that, even if it is not involved in reactivation for this task, REM does contribute to the consolidation of motor sequences.

Interestingly, our ERP results suggests that the brain can distinguish whether sounds are related to the left or right hand during REM. Indeed, previous research also suggests that discrimination of a stimulus' significance and semantic content may persist during REM (Bastuji & García-Larrea, 1999; Niiyama, Fujiwara, Satoh, & Hishikawa, 1994; Sallinen, Kaartinen, & Lyytinen, 1996; Takahara, Nittono, & Hori, 2006). However, these ERPs did not predict any form of consolidation which we

were able to measure, and the fact that stimuli can be processed during REM does not mean that this processing will lead to behavioural impacts.

2.5.2 TMR preferentially benefits the non-dominant hand

Prior studies of how TMR impacts upon the serial reaction time task have typically used only the left (non-dominant) hand (Cousins et al., 2014, 2016; Schönauer et al., 2014). To determine whether TMR differentially impacts upon dominant and non-dominant hands, we modified the task such that an equal number of responses were required from each hand. In the REM group, this did not appear to make a difference – TMR during this stage did not lead to significant benefits in either hand. However, we found an interesting difference in the SWS group. While there was a significant TMR effect when both hands were combined, only responses in the non-dominant hand showed a significant benefit when we analysed the hands separately. Responses in the dominant hand may have benefitted from TMR somewhat, but the difference between cued and uncued sequences did not reach significance in this hand.

If the non-dominant hand learned to a lesser extent in the first place our observations would fit with the literature indicating that TMR benefits weaker memories more than memories which were strongly learned (Cairney, Lindsay, Sobczak, Paller, & Gaskell, 2016; Drosopoulos, Schulze, Fischer, & Born, 2007; Schapiro, McDevitt, Rogers, Mednick, & Norman, 2018; Tambini et al., 2017). Indeed, we found that both raw RT and sequence-specific skill were worse in the left compared to the right hand, although this difference remained just shy of significance in the sequence-specific skill analysis. Our results thus largely fit with the idea that weaker memories particularly benefit from TMR. However, it may be the case that some other aspect of the way processing in the dominant and non-dominant hand differs underpins the observed consolidation bias.

Hand and finger movement representation in the primary motor cortex is significantly larger on the side contralateral to the dominant hand, which may be related to the greater motor skill often experienced in the preferred hand (Volkmann, Schnitzler, Witte, & Freund, 1998). It is possible that there is greater opportunity for gain in the non-dominant hand precisely because of its smaller motor skill repertoire. Of course, it is also possible that this result is related to the lateralised processing of hand movements rather than to a difference in handedness. The right motor cortex is primarily activated in response to contralateral (left) hand movements (Kim et al., 1993). In contrast, the left motor cortex is activated during both contralateral (right) and ipsilateral (left) hand movements, especially in right-handed participants. In other words, left-handed movements

produce activity in both hemispheres (Gut et al., 2007), and it is possible that this more widespread and balanced activation could be at the basis of the left hand benefit we find in our study. These ideas could be tested using a similar experiment with left handed participants. Such an experiment would help to clarify whether TMR is truly biased towards the non-dominant hand, or instead just towards the left hand.

2.5.3 Linear classifier with time domain features detects reactivation in SWS

While it is well established that TMR can facilitate consolidation, the question of whether this intervention truly triggers memory reactivation has attracted much attention in the last couple of years (see Lewis and Bendor, 2019; Schreiner and Staudigl, 2020 for reviews). A number of studies have now succeeded in demonstrating neural reactivation after TMR (Belal et al., 2018; Cairney et al., 2018; Murphy, Stickgold, Parr, Callahan, & Wamsley, 2018; Schreiner et al., 2018; Shanahan et al., 2018), using a variety of methods and measures. In this experiment, we developed a novel pipeline for classification of memory reactivation after TMR using EEG amplitude alone. Although we were able to reliably detect reactivation at above chance levels in SWS, there was no association between the level of detection and any measure of behavioural consolidation. This is in keeping with the findings of Belal and colleagues, who applied a different classification pipeline on the same task, but found no significant correlation with behaviour (Belal et al., 2018). Interestingly, however, some reports have identified correlations between detected reactivation and subsequent behavioural performance (Cairney et al., 2018; Schreiner et al., 2018; Shanahan et al., 2018). It is unclear whether this difference relates to the task in question or the specific classification pipeline.

In REM, our classification result was much more marginal. It is true that removal of an obvious outlier led to above-chance classification in the experimental night, and this was also significantly stronger than classification in the adaption night. However, the actual level of classification was still very low (mean of 51.6% correct). While this finding is encouraging and, in keeping with demonstration of differential ERPs for right and left hands in REM it suggests that TMR in this stage is eliciting some kind of response, it is not sufficient evidence to state that we can definitely detect reactivation in REM. Prior findings in our lab have suggested that TMR of this task is associated with REM-mediated changes in relevant brain areas (Cousins et al., 2016), also indicating some kind of reprocessing during REM sleep. However, the EEG in this sleep stage is extremely noisy, partially due to the many eye movements. We speculate that different features of the brain response may be needed to convincingly classify memory reactivation during REM. It will be interesting to see if

future studies using other pipelines are eventually able to detect REM reactivation in a more convincing manner.

It may initially seem difficult to reconcile the results from the ERP and classification analyses. The ERP analysis shows a difference between the response to cues associated with the left and right hand in the REM group, but no differences in the SWS group. In contrast, the classification analysis demonstrates reactivation of material in SWS but not REM. However, it is important to note that these analyses are not looking at the same thing. The ERP simply compares brain activity during sleep in response to cues associated with the right hand to activity in response to cues associated with the left hand. On the other hand, the classifier compares the neural signature in response to cues during wake with the neural signature in response to cues during sleep. It tries to find overlap between wake and sleep, with the goal of making a decision about which hand each particular cue may belong to. Thus, although the brain may be able to distinguish tones relating to the left and right hand during REM sleep, the activation during this stage may not resemble that during wake. In SWS, the differences between responses to cues associated with the left and right hand may be too small to be significant in an ERP analysis. Nevertheless, there appears to be meaningful overlap between cue-related brain activity during wake and during SWS. This overlap is such that the classifier is able to determine with above-chance accuracy whether a cue played during SWS is related to the left or the right hand.

2.5.4 Conclusion

In the current study, we demonstrate that the non-dominant hand benefits preferentially from TMR cued memory consolidation. This may be because the non-dominant hand is weaker in the first place, allowing more space for improvement, or because this hand places more bilateral demands on the brain, and may thus draw on more of the neural circuity that benefits from TMR. We also show that TMR in SWS, but not REM, leads to a consolidation benefit. Although our classifier was only able to reliably detect reactivation in SWS, not REM, our ERP results nevertheless demonstrate that the brain does process our stimuli during the latter sleep stage. These results suggest that, while the brain processes our cues during REM sleep, it may not necessarily be reactivating our procedural task in a manner that is conducive to consolidation during that sleep stage.

2.6 Supplements

2.6.1 Explicit Recall

For the explicit sequence knowledge scoring, individual items were only considered correct if they were in the correct position within the sequence, and if they were part of a segment that contains >2 correct items. This corresponded to the method used by Cousins et al. (2014).

To examine whether TMR leads to the overnight emergence of explicit knowledge, we conducted Wilcoxon signed-rank tests for each group separately. There was no difference between the cued and the uncued sequence in either the SWS or the REM group, see Table S1. In other words, TMR did not affect participants' explicit knowledge of the sequences in this experiment.

Table S1. *Explicit memory for each sequence, in number of correct items (± standard deviation).*

	REM sleep group *n = 15*	**SWS group** *n = 16*
Cued sequence	10.1 ± 2.8	9.8 ± 2.9
Uncued sequence	10.7 ± 2.4	9.3 ± 3.5
Significance value	p = 0.524	p = 0.837

2.6.2 Electrophysiology: Event-related potentials per night

Event-related potential (ERP) comparisons between the adaptation and experimental night were preprocessed and analysed in the same way as those concerning the left- and right-handed trials. In both the REM and the SWS group, ERPs to cues in the experimental night elicited a larger response than those in the adaptation night (see Figure S1a-b). However, this difference was not significant.

2.6.3 Time-frequency analysis

Time-frequency analysis used a 5-cycle frequency-dependent Hanning taper to obtain spectral power from 4-30 Hz in frequency steps of 0.5 Hz and time steps of 5 ms. These parameters were based on a previous study (Cairney et al., 2018). Averages across subject groups (REM or SWS) and

cue groups (left- versus right-handed or adaptation versus experimental night) were calculated as power change relative to a baseline window of -1 second until cue onset. To capture both slow and fast changes in the time-frequency representations, we looked at the entire time interval from cue onset (time 0) until another tone was played approximately 1.5 seconds later. Statistical analyses of time-frequency comparisons between cues were performed as paired-samples t-tests and corrected for multiple comparisons with FieldTrip's nonparametric cluster-based permutation method, using 1000 permutations. Results were considered significant at $p < 0.05$.

Time-frequency results comparing the nights in the REM group showed some slow activity approximately 250-500 ms after the cue (see Figure S1c-d). This activity seemed to occur in both nights, and a cluster permutation test confirmed that the difference between the nights was not significant. Results in the SWS group revealed early (around 500 ms after the cue) theta-band activity, and a weak fast spindle-band response to experimental cues around 1.2 seconds after cue onset (see Figure S1e-f). Again, these responses were similar between the nights, and statistical analyses showed that the difference was not significant.

When looking at the left- and right-handed cues during the experimental night, we found that the REM group again showed early theta or even delta activity at around 250-500 ms after the cue (see Figure S2a-b). In response to the left-handed cues there also seemed to be some fast spindle-band activity starting around 500 ms after the cue, while this was not the case when looking at the right-handed cues. However, statistical analyses did not show a significant difference between the EEG responses to these different cues. This remained the case when the left and right hemisphere were analysed separately.

In the SWS group, both time-frequency results showed theta-band activity around 500 ms after the cue (see Figure S2c-d). This activity even reached the alpha band in response to right-handed cues. In contrast, there seemed to be a slightly more pronounced spindle-band response around 1.1-1.2 seconds after left-handed cues. Statistical analyses comparing these responses did not reveal any significant differences. The two hemispheres showed remarkably similar results, and differences in the time-frequency responses to left- and right-handed cues thus remained non-significant when we analysed the hemispheres separately.

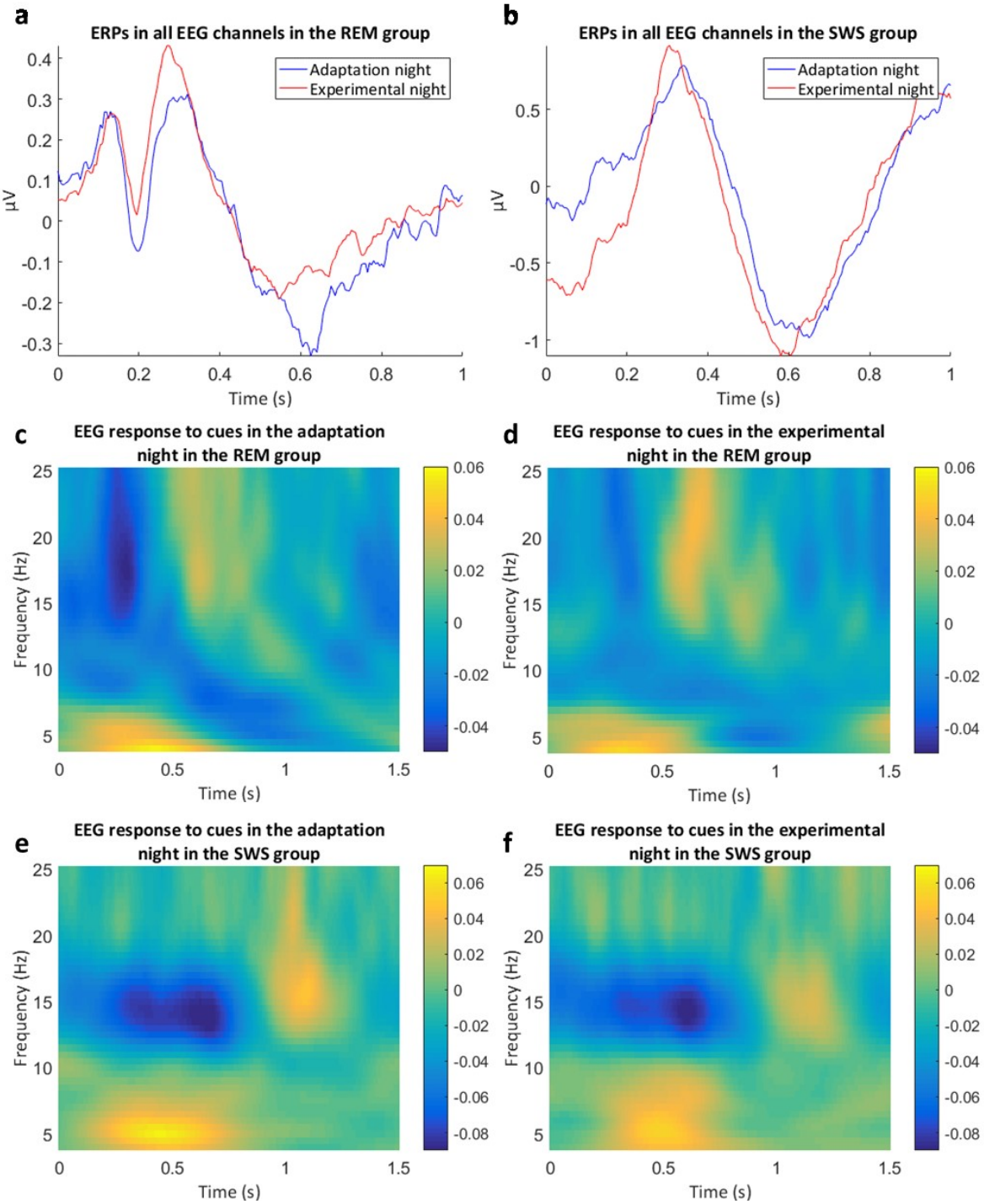


Figure S1. *Electrophysiological results comparing the adaptation and experimental night in all EEG channels.* ***a)*** *ERPs in all EEG channels in the REM group.* ***b)*** *ERPs in all EEG channels in the SWS group.* ***c)*** *Time-frequency representation of the EEG response to cues in the adaptation night in the REM group.* ***d)*** *Time-frequency representation of the EEG response to cues in the experimental night in the REM group.* ***e)*** *Time-frequency representation of the EEG response to cues in the adaptation night in the SWS group.* ***f)*** *Time-frequency representation of the EEG response to cues in the experimental night in the SWS group. Legends in the time-frequency plots represent relative change from baseline.*

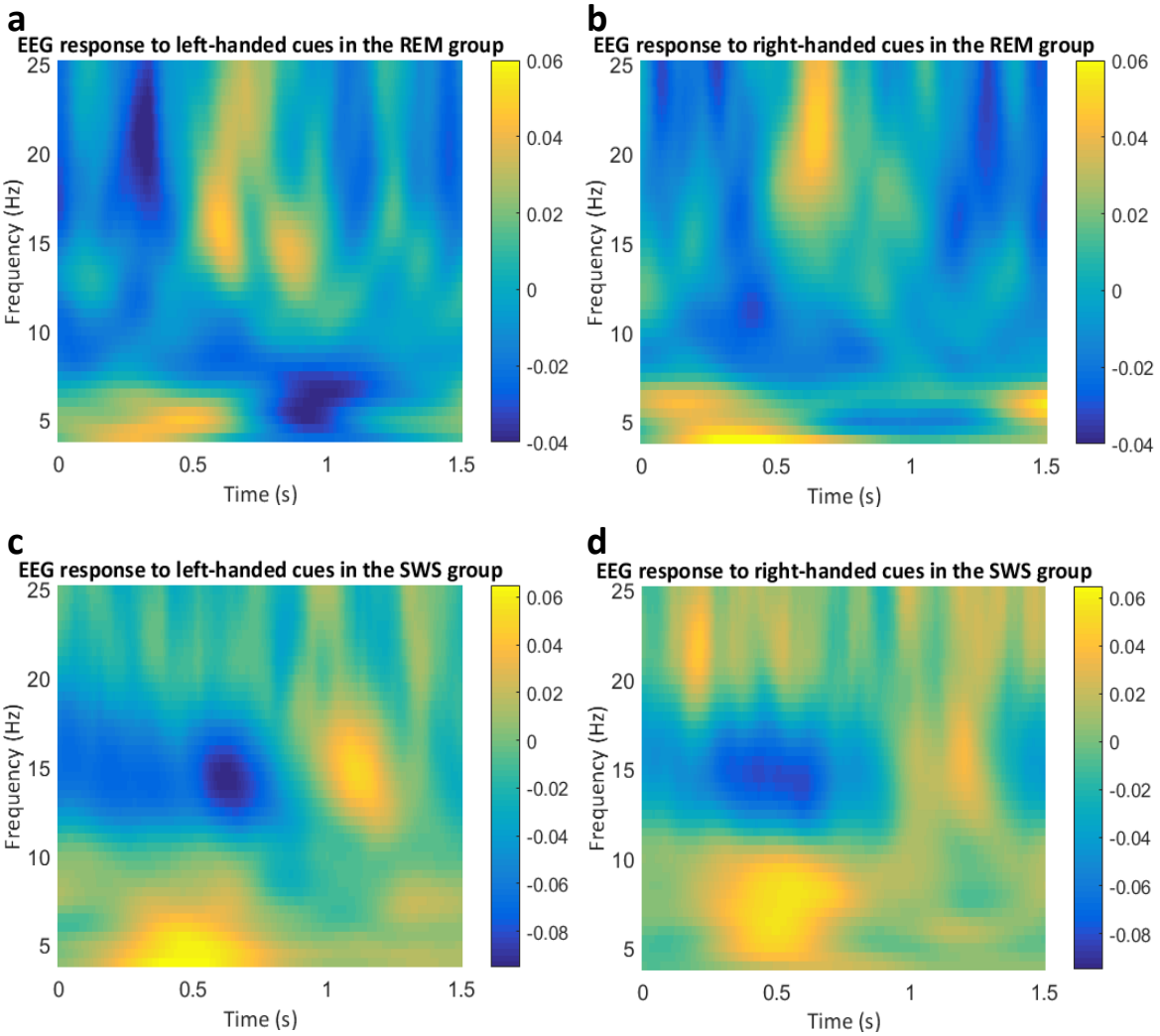


Figure S2. *Time-frequency results comparing cues related to the left and right hand, in all EEG channels. Legends represent relative change from baseline.* ***a)*** *Time-frequency representation of the EEG response to cues associated with the left hand in the REM group. b****)*** *Time-frequency representation of the EEG response to cues associated with the right hand in the REM group.* ***c)*** *Time-frequency representation of the EEG response to cues associated with the left hand in the SWS group. d****)*** *Time-frequency representation of the EEG response to cues associated with the right hand in the SWS group.*

CHAPTER 3

The effect of targeted memory reactivation in REM and SWS on remote associations

3.1 Abstract

Sleep has been shown to improve the ability to make indirect associations between remotely related items that were not directly learned. However, the involvement of memory reactivation and the roles of different sleep stages remain unclear. Experiment 1 examines the impact of targeted memory reactivation (TMR) of an associative memory task during two sleep stages: rapid eye movement (REM) and slow-wave sleep (SWS). Thirty-two participants (16 each in the REM and SWS groups) learned to match 40 sounds to semantically related scenes. Participants then learned to independently associate two faces with each scene-sound pair. Following learning, participants' sleep was monitored and half of the sounds were replayed to them either in REM or SWS. In the morning and in a two-week follow-up, we tested face-scene associations, which had been learned directly, and face-face associations, which had to be inferred. In Experiment 1, the REM group performed significantly higher on cued compared to uncued items in the remote (face-face) associations test. However, two independent replication experiments of the REM group with new participants (n = 24 and n = 20, respectively) did not show the same effect. Moreover, although electrophysiological results in the original REM group suggest that memory processing occurs during REM sleep, and that we can trigger it using learned sounds, the replications do not show the same electrophysiological pattern. Although there may be individual differences in the receptiveness to TMR, we must conclude that TMR during REM sleep does not reliably strengthen indirect associations in this task.

3.2 Introduction

A wealth of studies has outlined the benefits of sleep on memory (Diekelmann & Born, 2010; Rasch & Born, 2013). For instance, sleep has been known to improve performance on tasks that measure the association between images and spatial locations (Rasch et al., 2007; Rudoy et al., 2009) or words in different languages (Göldi et al., 2019; Schreiner & Rasch, 2015). However, there is now an increasing amount of evidence that sleep not only strengthens directly associated memories, but also aids the formation of novel or indirect associations. With these indirect associations, we are able to connect experiences that are not directly related, but which may share common characteristics or components.

Models of declarative memory have suggested that our memories are not stored in isolation, but instead exist in interconnected networks of related experiences (Eichenbaum, 2004; McClelland et

al., 1995; O'Reilly & Rudy, 2001). This 'relational memory' may be incidental, where components of an experience coincidentally correspond to another experience (Konkel & Cohen, 2009). Consequently, these indirect associations allow us to generalise across experiences and flexibly make use of memories in situations that are new to us (Eichenbaum, 2004). In other words, the ability to make remote associations and inferences is key to adapting to new situations, and sleep has been shown to improve this ability (Alger & Payne, 2016; Lau et al., 2010).

Two studies have investigated the effects of sleep and wakefulness on direct and remote associations (Alger & Payne, 2016; Lau et al., 2010). They used a previously-developed associative inference task (Bunsey & Eichenbaum, 1996; Preston et al., 2004). In the task, the objective was to learn pairs of faces and objects, where the objects were always paired with two separate faces. Importantly, the faces were never learned together, and their relationship had to be inferred through the shared object. After a period spent either asleep or awake, these remote face-face associations were tested, followed by the learned face-object associations. Lau and colleagues first showed that a nap improved accuracy on both learned and remote associations compared to wakefulness (2010), and this was later replicated (Alger & Payne, 2016). Conflictingly, these two studies differed in their findings regarding the stage of sleep that was related to increased inference ability. In one, this ability was correlated with the duration of slow-wave sleep (SWS) (Lau et al., 2010), although participants who entered rapid eye movement (REM) sleep were not included in the analyses. In contrast, the other study found that the percentage of REM sleep was negatively related to performance on the learned associations, but positively correlated with accuracy on the remote associations (Alger & Payne, 2016). It is worth noting that this second study used both emotional and neutral stimuli, which may be relevant given that REM sleep has been implicated in emotional memory (Hutchison & Rathore, 2015). However, the relationships between task performance and REM sleep in this experiment were both present in the neutral, not the emotional, stimuli (Alger & Payne, 2016).

Sleep is thought to improve relational memory through memory reactivation, whereby neural traces encoded during the task are reprocessed during sleep. A model outlined by Lewis and colleagues (2018) has proposed complementing roles for SWS or non-REM and REM in this process. Specifically, the model describes the role of non-REM sleep to be the extraction of common features from overlapping memories. When these memories are reactivated, their overlapping parts are strengthened and this forms a schematic representation of related memories. This is called information overlap to abstract (iOtA) and specifies the non-REM part of the model (Lewis &

Durrant, 2011). The broader form of this model, called BiOtA, posits that subsequent concurrent reactivation of existing schemas during REM sleep can lead to detection of a shared underlying structure.

No studies to date have directly examined the effect of memory reactivation on the associative inference task. An interesting technique to evaluate this link is targeted memory reactivation (TMR), whereby task items are paired with sensory stimuli, and subsequent re-presentation of these stimuli during sleep leads to memory reactivation of the task items (Belal et al., 2018; Schreiner et al., 2018; Shanahan et al., 2018) and increased performance on those cued items after sleep (e.g. Cousins, El-Deredy, Parkes, Hennies, & Lewis, 2014; Rudoy et al., 2009; Schreiner & Rasch, 2015). This method could also more directly test the involvement of different sleep stages in the found effect of sleep on associative memory. The nap studies mentioned above have indicated that REM and SWS may be involved, and the BiOtA model has outlined complementary roles for these stages in associative memory. Taking a closer look at these roles, using TMR, will allow us to examine some of the ways that sleep may influence relational memory.

The current study uses a between-participant design to investigate the differential impact of TMR of an associative memory task during REM or SWS. Participants learned to match 40 sounds to semantically related scenes. Then, they learned to independently associate two faces with each scene-sound pair. This design was very similar to previous studies, the differences being that we used scenes rather than objects, and we added sounds to allow for TMR. Following learning, participants' sleep was monitored and half of the sounds were replayed to them either in REM or in SWS. In the morning and in a two-week follow-up, learned face-scene associations and remote face-face associations were tested. Furthermore, we conducted two replications of the REM group (Experiments 2 and 3). We hypothesised, based on the extensive literature that outlines the importance of SWS on declarative memory, that learned associations would be strengthened by reactivation of the sounds during SWS. In other words, we expected that participants would perform more accurately on those learned associations which corresponded to the sounds that had been cued in SWS, compared to those that had not been cued. We expected this to be the case both immediately after sleep as well as during the two-week follow-up. The remote associations, on the other hand, were not directly learned but shared an underlying structure (i.e., these faces were paired with the same scene). Following the BiOtA model, we predicted that these remote associations would be strengthened by reactivation during REM sleep. Thus, we hypothesised that accuracy would be higher on those remote associations which corresponded to the sounds that had

been cued in REM, compared to those that had not been cued. We expected this difference to be present the morning after sleep and at the two-week follow-up.

Although our main analyses focus on accuracy, we were also interested in the effect of TMR on reaction times, because this could indicate a more implicit learning of the scene-face combinations. If this implicit learning indeed took place, it could have led to faster (lower) reaction times on cued items compared to uncued items. Note that we did not necessarily expect these differences in reaction times to be present, but we were interested in exploring them. Finally, we hypothesised that we would find indications of memory reactivation in the electrophysiological results. During sleep, we played TMR cues as well as control sounds that had been heard twice during wake and had not been coupled with a face. Thus, we expected to find differences between our TMR cues and these control sounds without a strong memory component in our analyses of event-related potentials and time-frequency data, in both the REM and the SWS group.

3.3 Experiment 1: Materials and Methods

3.3.1 Participants

For this study, a total of 54 participants were recruited. Six participants were excluded from analyses, five due to having <5 hours of sleep, and one because of experimenter error regarding the tasks. The final sample, then, was 48 participants (23 females, aged 19-30 years). These were allocated to one of three groups of 16 participants each: the Control group with TMR during wake (7 females, mean age 22.3 ± 2.9), the REM group (8 females, mean age 22.4 ± 2.6), and the SWS group (8 females, mean age 23.9 ± 3.5). All participants were required to have normal or corrected-to-normal vision and normal hearing, be non-smokers, and have no history of psychological, neurological, or sleep disorders. In their responses to a pre-screening questionnaire, they also reported no use of any psychoactive medications, a lack of regular night work, and generally regular sleep. All participants had native or near-native levels of English and they were required to abstain from alcohol, caffeine, and napping for 24 hours prior to the overnight part of the experiment.

This study was approved by the School of Psychology, Cardiff University Research Ethics Committee, as well as the University of Manchester Research Ethics Committee, and all participants gave written informed consent.

3.3.2 Purpose

The purpose of this study was to test whether TMR during REM sleep and SWS improved participants' ability to remember learned face-scene-sound associations, and make remote face-face associations. To that end, subjects learned to match scenes and sounds during two short day-time sessions. Then, participants returned to spend a night in the lab, during which they learned which faces belonged with which scene-sound pair. They were tested on these learned associations before and after sleep. In the morning they were further tested on the remote associations, i.e. the faces which were not learned together but had been paired with the same scene. The task structure can be found in Figure 3.1. Note that a Control group with TMR during wake was added to check whether effects of TMR we found were specific to sleep. Such a control group was not added in the experiment in Chapter 2, because this had already been done in a previous study (Cousins et al., 2014).

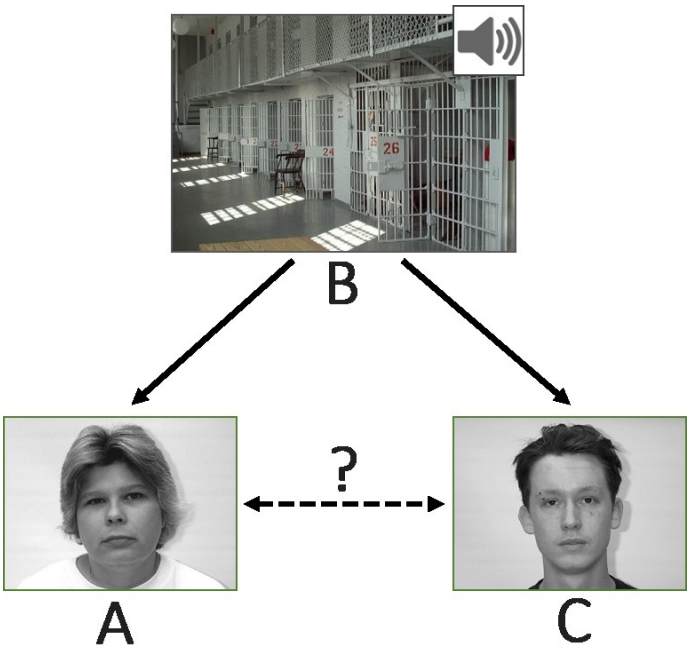


Figure 3.1. *Task structure. Participants first learned two independent face-scene associations (A-B and B-C pairs). They were later surprised by a test probing face-face associations (A-C pairs).*

3.3.3 Stimuli

Stimuli consisted of 60 colour images of different scenes and 60 matching 2-second long sounds all taken from the internet. It is important to note that the sounds were semantically related to the

scenes, to make these combinations easy to remember. Examples of the scene-sound pairs are: a picture of a back garden with the sound of chirping birds, a picture of a bathroom with the sound of a toilet flushing, and a picture of rush hour traffic with the sound of honking cars. For each participant, 40 of the scene-sound pairs were randomly chosen to be thoroughly learned. The remaining 20 were left as control sounds. These control sounds were heard only twice during the learning sessions which precede the overnight part of the experiment. During the night, as participants slept, the control sounds were played (in addition to the sounds meant for TMR) to allow comparison of the brain's response to memory-related cues (TMR sounds) and cues without a strong memory component (control sounds).

Participants further learned to match a male and a female face (separately) to each of the scene-sound pairs. In other words, they independently learned 40 female and 40 male faces, which were taken from a face database that includes faces of all ages (Minear & Park, 2004).

Images and tasks were presented on a computer monitor with 1024 x 768 pixel resolution and using Matlab (R2015a, The Mathworks Inc., Natick, MA) and Cogent 2000 (Functional Imaging Laboratory, Institute for Cognitive Neuroscience, University College London). Sounds were played through noise-cancelling headphones (Sony MDR-ZX110NA) during the tasks and through speakers (Dell A225) during sleep.

3.3.4 Experimental protocol

Before the overnight part of the experiment, participants performed two learning sessions; session 1 took place 1-3 days before the overnight, and session 2 took place on the morning of the overnight. The goal of these sessions was to make sure participants learned the combinations of scenes and sounds used throughout the experiment to criteria. As the sounds were to be played while participants were sleeping, and since they were intended to evoke the memory of the associated item in the task, it was important to ensure that these scene-sound combinations were well-learned. To this end, participants completed the following tasks: simple viewing (SV), multiple choice (MC), and free recall (FR) in the order of SV → MC → FR → MC → FR (session 1) and FR → SV → MC → FR (session 2). During SV, each of the 60 scenes was presented on the screen for 3 seconds, together with its corresponding sound. A 500 ms fixation cross separated the trials. The MC and FR tasks were done with a subset of 40 scene-sound pairs, which were randomly selected for each participant but kept constant for that participant. During the MC task, participants were shown 6 randomly selected scenes. These scenes stayed on the screen until the participant

responded, and the sound corresponding to one of these scenes was played after 500 ms. Their task was to indicate, using numbers on the keyboard, which scene corresponded to the sound they just heard. Feedback was given in the form of a green check mark if correct, and a red cross with the correct answer underneath if incorrect. Finally, the FR task required participants to listen to their 40 sounds in succession. After hearing each sound, they wrote down a short description of the matching scene on a sheet of paper, and pressed space bar to move on to the next sound. After hearing all the sounds, these were played again in the same order. This time, the scenes were also pictured, allowing participants to check whether their written description matched the picture. All participants had reached 100% accuracy on the free recall task by the end of the second session.

On the evening of the day of session 2, participants returned to the sleep laboratory approximately 3 hours before their normal bedtime. They were asked to change into their sleepwear, and were fitted for polysomnography (see section 3.3.6 for details). After completion of the Karolinska Sleepiness Scale (KSS; Åkerstedt & Gillberg, 1990) and the Stanford Sleepiness Scale (SSS; Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973), the encoding task was started. This task required participants to learn 40 female and 40 male faces that each matched with one of the scene-sound pairs. Scenes and faces were randomly paired and presented in a randomised order, but we ensured that the two appearances of the same scene were separated by at least 9 other scenes. The position (left or right) of the face in relation to the scene was also randomised. Each face-scene pair was displayed for 2.5 seconds, during which time the corresponding sound was also played. A 500 ms inter-trial fixation cross followed each presentation. The first three trials in the encoding phase were dummy trials (with scenes and sounds that had not been presented before); this was to allow the participants to get used to the way the scenes and faces were presented. Encoding was immediately followed by a test, to see how well the face-scene-sound associations had been learned. As with encoding, the first three trials were dummies, and trials were separated by a 500ms fixation cross. The test was multiple choice – a scene was presented on the top half of the screen, and four faces with corresponding letters (to press on the keyboard) were shown at the bottom. The sound that matched with the scene was also played, and participants could choose to hear this sound again by pressing ‘N’ on the keyboard. If the participant had 66% or more correct on this test (53 or more out of 80 items), the task ended. If not, another learning and test round followed, with a maximum of 3 learning rounds.

After completing the face-learning task, participants performed a 2-back task (adapted from Kane, Conway, Miura, & Colflesh, 2007). This task took 45 minutes in total, divided into three rounds of

15 minutes. Each trial in this task consisted of the presentation of one of eight phonologically distinct letters (B, F, H, K, M, Q, R, X). The letter was displayed in white in the middle of a black screen for 500 ms, and was followed by a black screen inter-trial interval of 2500 ms. Participants' objective in the task was to press one button ('Yes') if the letter currently on the screen matched the letter which appeared two items ago (e.g. M-k-m), and a different button ('No') if these letters did not match. Letters appeared randomly in upper or lower case, to prevent recognition based only on perceptual features. Instructions were to respond as quickly and accurately as possible, and responses were accepted as soon as a letter appeared on the screen and until the end of the inter-trial interval. Each round of 15 minutes was further divided into six blocks of 48 trials each. Every letter appeared six times in each block, once as a target and five times as a foil. Participants were told to keep focusing on the task and try their best. Feedback on accuracy was given after each block, and participants were encouraged to maintain (if 100%) or improve their performance in the next block. In the Control group, TMR of the learned sounds took place during this 2-back task, to prevent active listening or rehearsal of the sounds. The SWS and REM groups also performed the 2-back task, and were asked to wear the headphones ('for noise cancelling purposes'), but no sounds were played for these two groups during the task.

Subsequently, participants went to sleep. Targeted memory reactivation took place as detailed in section 3.3.5. After approximately 7-8 hours of sleep, participants were woken up, though care was taken not to wake them from SWS or REM. Their electrodes were removed, and they were given the opportunity to take a shower. Then, after filling in a sleep quality questionnaire (adapted from Görtelmeyer, 1985) and the KSS and SSS again, the morning tests were started. The first test was the inference test, wherein participants were asked which people (faces) were previously shown with the same scene. This information was not learned or explicitly pointed out to participants the previous evening and therefore had to be inferred. A female face was presented at the top of the screen, and three male faces were presented at the bottom. Participants indicated which faces had been paired with the same scene by pressing a corresponding button on the keyboard. This button press initiated the following trial, after a 500 ms inter-trial fixation cross. The inference test was followed by the recognition test, which asked which face belonged to which scene as the evening before. Participants were then free to leave. A visual description of the experimental protocol is found in Figure 3.2.

3.3.5 Targeted memory reactivation

After the encoding task in the evening, the results files for each participant were copied and used to divide the items into three different categories: the control list, the cued list, and the uncued list. The control list was simply those 20 items that were not thoroughly learned during the learning sessions and were not paired with a face. The cued and uncued list each consisted of half of the items seen during the encoding task, thus 20 items each. They were matched to each other in terms of accuracy and reaction time, which ensured that performance on the cued and uncued items was similar before sleep. The control list and cued list were added together, and these 40 sounds were played during the TMR. In total, a maximum of 600 sounds were played for each participant, i.e. each sound 15 times. The average number of sounds played in the Control group was 495.3 (each sound 12.4 times), in the SWS group 585.3 (each sound 14.6 times), and in the REM group 582.9 (each sound 14.6 times).

As mentioned above, for the Control group, TMR took place while participants were awake and performing the 2-back task. The other participants' sleep was monitored, and after they entered stable SWS or REM (depending on the group) the reactivation was started. Reactivation was paused immediately when participants showed signs of arousal, or when they left the relevant sleep stage. In the case of the Control group, reactivation was paused after each 15-minute block of the n-back task. TMR was continued until all 600 sounds were played, or until it was time for the participant to wake up – whichever criterion was met first.

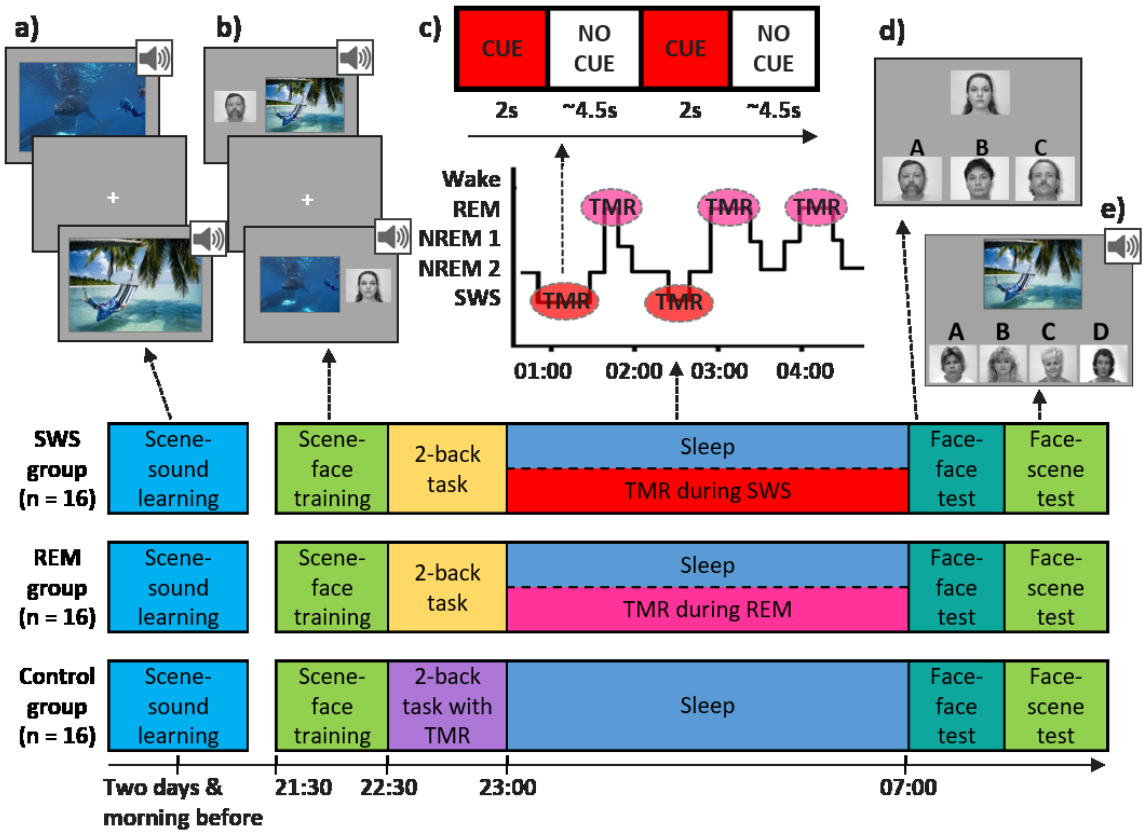


Figure 3.2. *Experimental procedures.* ***a)*** *Scenes appeared together with a unique sound. Other tasks during the scene-sound learning included a multiple choice task where participants heard a sound and had to pick the correct picture from six options, and a free recall task where participants heard a sound and had to write a description of the scene (both not pictured).* ***b)*** *Forty scenes/sounds were paired with a female and a male face. Each face-scene pairing was learned separately. Training was immediately followed by testing as described in e).* ***c)*** *Twenty of the sounds that had been paired with faces, interleaved with twenty control sounds, were played to the participant either during SWS (SWS group), REM (REM group), or the 2-back task (Control group). Presentation of each sound (CUE) was followed by approximately 4.5 seconds of silence (NO CUE). The hypnogram shows TMR during SWS and REM.* ***d)*** *In the morning, participants were first tested on the remote associations, by picking which faces had been paired with the same scene.* ***e)*** *Then, they were retested on the learned associations by matching which face had been paired with which scene.*

3.3.6 Polysomnography (PSG) data acquisition and analysis

The participants for this study were collected in two batches. For the majority of the Control group, 17 electrodes were placed on the scalp and face of the participants following the 10-20 system. On the scalp, these were at 10 standard locations: F3, F4, C3, C4, P3, P4, P7, P8, O1, and O2, and they were referenced to the mean of the left and right mastoid electrodes. Further electrodes used were the left and right EOG, two EMG electrodes on the chin, and the ground electrode on the forehead. The impedance was <5kΩ for scalp electrodes, and <10kΩ for face electrodes. Recordings were made with an Embla N7000 amplifier and RemLogic 1.1 PSG Software (Natus Medical Incorporated) and sampled at 200 Hz.

Recordings for the REM and SWS groups, and 4 participants within the Control group, were made with BrainVision Recorder and BrainAmp MR Plus amplifiers (Brain Products GmbH, Gilching, Germany) and sampled at 500 Hz. Electrodes and impedance were the same, but the standard locations Fz, Cz, and Pz were added to enhance electrophysiological analyses. PSG recordings were manually scored by two trained sleep scorers according to the standard AASM criteria (Berry et al., 2015). Both scorers were blind to the periods when sounds were reactivated.

3.3.7 Additional tasks

The second batch of participants, those that were collected using the BrainVision system, performed three additional tasks after completion of the experimental protocol above. On the morning after sleeping in the lab, they completed the Cambridge Face Memory Test (Duchaine & Nakayama, 2006), to allow examination of whether general face recognition skill would interact with TMR effects. Furthermore, these participants were asked to come back two weeks after sleeping in the lab, to perform the inference and recognition tests again. All participants in the SWS and REM groups returned for this follow-up, on average 13.6 days (range 9-16 days) after their monitored night. There was no difference in the follow-up time between the groups ($t(30) = 0$; $p = 1$), as assessed with an independent samples t-test.

3.3.8 Behavioural data analysis

Performance was assessed by measuring accuracy (proportion of correct answers) in the recognition and inference tasks. The recognition test had three time points: pre-sleep, post-sleep, and two-week follow-up. The inference test was only conducted post-sleep and at the two-week follow-up, because presence of this test pre-sleep would have alerted participants to the purpose of the study. The two-week follow-up was not administered in the majority of the control group, and therefore we did not take this time point into account in any analyses looking at the control group. Accuracy in the two tasks was investigated with analysis of variance (ANOVA) to evaluate the effects of time and TMR, separately for each group. ANOVAs were conducted in R (version 3.6.3, R Core Team, 2020) with the “afex” package (Singmann et al., 2020). This package automatically applies the Greenhouse-Geisser correction for sphericity when Mauchly’s test of sphericity is violated. Whenever this was the case, epsilon-corrected degrees of freedom are given in the text. As planned comparisons, two paired t-tests were conducted on the face-face task data from the REM group, to compare performance on cued and uncued items at both the post-sleep and two-week time points. Planned paired t-tests were also conducted on the face-scene data from the SWS

group, to compare performance on cued and uncued items at the post-sleep and two-week time points. All statistical tests were 2-tailed and considered significant for $p < 0.05$. We corrected for multiple comparisons with the false discovery rate (FDR) method, which takes into account the expected proportion of falsely rejected hypotheses (Benjamini & Hochberg, 1995). We included measures of effect size: generalised eta squared (η^2_G) for ANOVA as calculated with the "afex" R package (Bakeman, 2005; Lakens, 2013; Olejnik & Algina, 2003; Singmann et al., 2020), and Hedges' g for t-tests as calculated with the "effsize" R package (Hedges, 1982; Lakens, 2013; Torchiano, 2020).

3.3.9 EEG data analysis

The EEG data collected during sleep were analysed using MATLAB (version R2016b) and the FieldTrip Toolbox (version 20/08/2019; Oostenveld, Fries, Maris, & Schoffelen, 2011). The continuous sleep data were preprocessed by filtering between 0.1 Hz and 30 Hz, and subsequently segmented into epochs starting from 1 second before the onset of a cue until 4 seconds after. Artifacts were rejected in a multi-step procedure. First, trials were re-segmented into slightly smaller trials of -0.5 and +3 seconds around cue onset. A trial was considered an outlier for a given channel if it was more than two standard deviations from the mean on amplitude or variance. If more than 25% of channels (i.e. more than 3 channels) showed an outlier in a given trial, that trial was excluded. This resulted in the removal of on average 12.5% of trials in the SWS group, and 10.2% of trials in the REM group. The remaining trials which contained outliers in <25% of channels were interpolated based on triangulation of neighbouring channels. To remove eye movements in the data from the REM group, independent component analysis (ICA) was used. Independent components identified in the ICA were correlated with the signal from the EOG channels and significantly correlated components (corrected for multiple comparisons) were removed. Finally, all channels were manually inspected after these procedures, and channels that still looked noisy were interpolated based on their neighbours. Analyses focused on the difference in brain responses to the experimental (memory-related) versus control cues.

Event-related potentials (ERPs) were analysed time-locked to TMR cue start, and baseline-corrected to a baseline window of -1 second until cue onset. This long baseline window was chosen because of the low-frequency nature of SWS, and kept the same for consistency in the REM analyses. ERPs were expected to occur in the frontal and central channels during the 2-second cue. Therefore, statistical analyses used averaged data of channels C3, Cz, C4, F3, Fz, and F4 from cue onset until cue end.

Due to the pronounced difference in electrophysiology between SWS and REM, comparisons between ERP responses to experimental and control sounds were done separately for the SWS and REM groups. They were performed as paired-samples t-tests and corrected for multiple comparisons using FieldTrip's nonparametric cluster-based permutation method, using 1000 permutations. Results were considered significant at $p < 0.05$.

Time-frequency analysis used a 5-cycle frequency-dependent Hanning taper to obtain spectral power from 4-30 Hz in frequency steps of 0.5 Hz and time steps of 5 ms. Averages across subject groups (REM or SWS) and sound groups (experimental or control) were calculated as power change relative to a baseline window of -1 second until cue onset. Averages of channels C3, Cz, C4, P3, Pz, and P4 were taken, as power changes were considered most likely in these electrodes. Statistical analyses of time-frequency comparisons between experimental and control sounds were performed as paired-samples t-tests and corrected for multiple comparisons with FieldTrip's nonparametric cluster-based permutation method, using 1000 permutations. Results were considered significant at $p < .05$.

3.4 Experiment 1: Results

3.4.1 Sleep data

Sleep data of all groups can be found in Table 3.1. Results from ANOVAs with between-subject factor group and the different sleep stages as dependent variables showed that there was no effect of group on time spent in any of the sleep stages or total sleep time (lowest $p = 0.128$, uncorrected).

Table 3.1. *Average minutes spent in sleep stages (± standard deviation), and p-values for the group difference.*

	SWS group	REM group	Control group	Significance values of group difference
Stage 1	36.97 ± 19.34	41.56 ± 23.28	40.53 ± 17.28	p = 0.796
Stage 2	238.53 ± 30.11	228.09 ± 32.38	234.25 ± 37.95	p = 0.680
SWS	94.31 ± 27.32	88.09 ± 29.55	90.03 ± 42.20	p = 0.867
REM	90.13 ± 19.06	109.78 ± 35.04	105.09 ± 27.60	p = 0.128
Total sleep time	459.94 ± 25.55	467.53 ± 41.43	469.90 ± 43.66	p = 0.739
Wake after sleep onset	16.46 ± 13.13	18.06 ± 20.85	16.81 ± 25.44	p = 0.974

3.4.2 Learned associations (face – scene)

We separately tested the effect of TMR in each group (SWS, REM, Control) on the learned associations. Thus, we conducted three repeated measures ANOVAs with within-subject factors time (evening, morning, and two weeks later) and reactivation (cued or uncued). Results showed a main effect of time in both the SWS ($F(1.87,28.07) = 22.60$; $p < 0.001$, $\eta^2_G = 0.112$) and the REM groups ($F(1.53,22.91) = 5.90$; $p = 0.013$, $\eta^2_G = 0.037$), but not in the Control group ($F(1,15) = 0.00$; $p = 0.999$). This result seemed largely to be caused by decreased performance at the two-week follow-up (see Figure 3.3). Critically, there was no main effect of reactivation or any interaction with reactivation (lowest $p = 0.129$, in the Control group). Follow-up planned paired t-tests showed that participants in the SWS group did not perform better on cued compared to uncued items directly after sleep ($t(15) = 0.18$; $p = 0.857$), nor in the two-week follow-up ($t(15) = -0.49$; $p = 0.630$).

We further analysed whether time spent in REM, SWS, or Stage 2 correlated with the benefit of TMR on the learned associations (i.e. the number of correct items in the cued condition minus the number of correct items in the uncued condition). TMR benefit in the morning or at the two-week follow-up did not significantly correlate with time spent in any of these sleep stages, in any of the groups (lowest $p = 0.306$, after correction for multiple comparisons with the FDR method). Finally, due to individual sleep characteristics, there was some variability in the amount of TMR sounds played during the night, which may have affected TMR benefit. In the REM group, there was a trend towards a negative association between the amount of cues played and TMR benefit at the two-week follow-up face-scene test ($r = -0.518$; $p = 0.079$, corrected). There were no other correlations that approached significance (lowest $p = 0.238$, corrected).

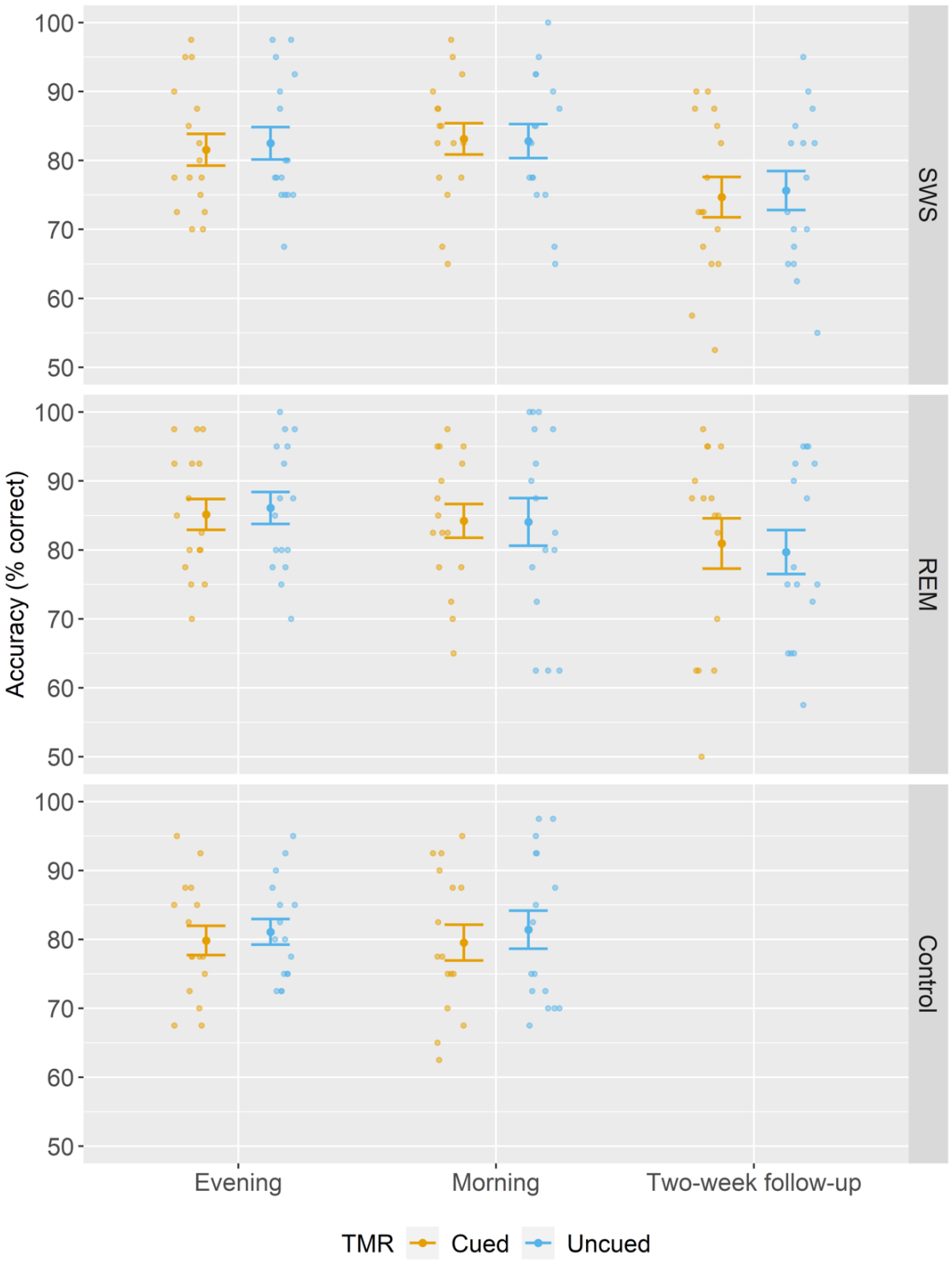


Figure 3.3. *Accuracy on the learned associations in the SWS, REM, and Control groups at different time points. Error bars represent 1 standard error of the mean (SEM).*

3.4.3 Remote associations (face – face)

To analyse the remote associations, we again separately tested the effect of TMR in each group (SWS, REM, Control). We thus conducted three separate repeated measures ANOVAs with within-participant factors time (morning and two-week follow-up) and reactivation (cued and uncued). In the SWS group, there was a trend for an effect of time ($F(1,15) = 3.97$; $p = 0.065$, $\eta^2_G = 0.028$), with decreased accuracy at the two-week follow-up as compared to the post-sleep test (see Figure 3.4). There was no effect of reactivation in the SWS or the Control group (lowest $p = 0.594$). In contrast, in the REM group, the main effect of time was significant ($F(1,15) = 7.04$; $p = 0.018$, $\eta^2_G = 0.045$). Furthermore, there was a main effect of reactivation in the REM group ($F(1,15) = 7.53$; $p = 0.015$, $\eta^2_G = 0.038$). Follow-up planned paired t-tests showed that participants in the REM group performed better on cued compared to uncued items in the two-week follow-up ($t(15) = 2.76$; $p = 0.029$, Hedges' $g = 0.451$, FDR corrected for multiple comparisons), but not in the test directly after sleep ($t(15) = 1.57$; $p = 0.138$, corrected).

Like the learned face-scene associations, the benefit of TMR (i.e. the number of correct items in the cued condition minus the number of correct items in the uncued condition) in these remote associations, either in the morning or in the two-week follow-up, did not significantly correlate with time spent in any particular sleep stage in any group (lowest $p = 0.621$, corrected). TMR benefit also did not correlate with the amount of TMR sounds played, in any group or at any time point (lowest $p = 0.766$, corrected).

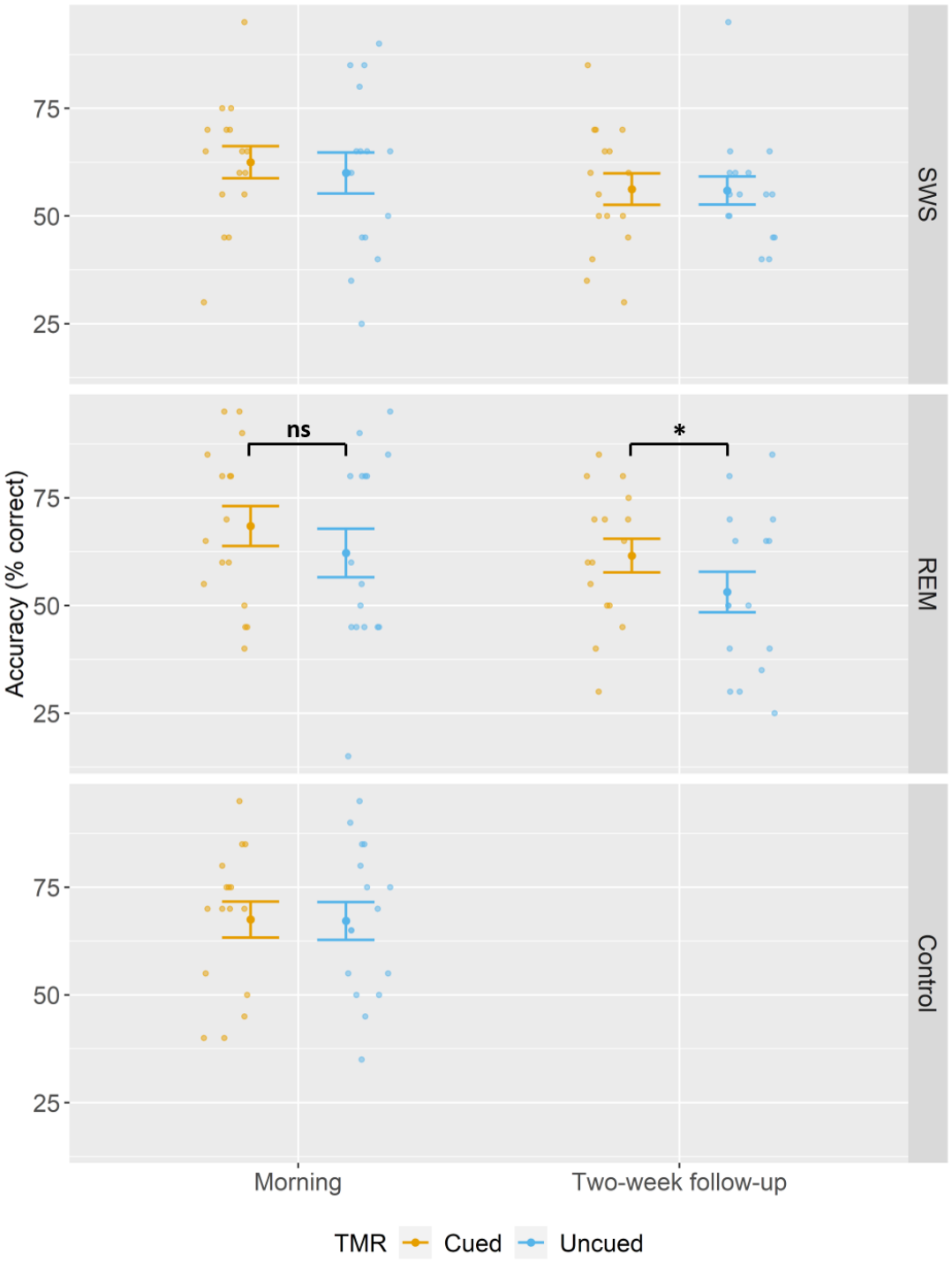
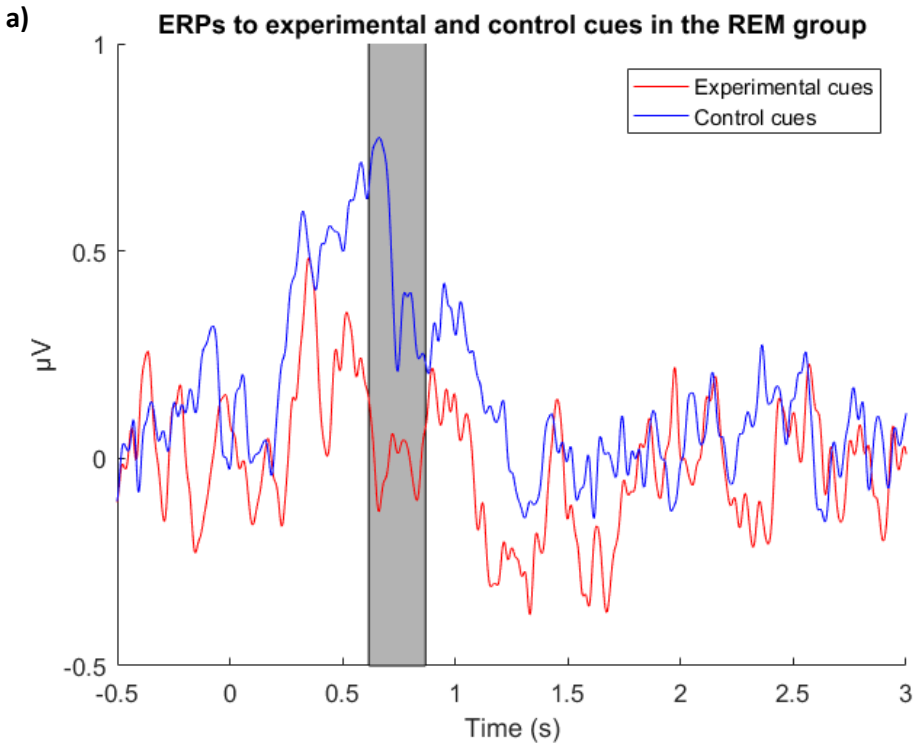


Figure 3.4. *Accuracy on the remote associations in the SWS, REM, and Control groups at different time points. * p = 0.029 (FDR corrected); ns = not significant, p = 0.138. Error bars represent 1 standard error of the mean (SEM).*

3.4.4 Event-related potential analysis

Due to the differential oscillatory dynamics of REM and SWS, ERPs in response to experimental and control cues were analysed separately for the SWS and REM groups. In the REM group, there was a strong k-complex-like response to the control cues that was largely absent in response to the memory-related (experimental) cues. This difference was significant at $p < 0.05$ between 0.61 and 0.87 seconds after cue onset (see Figure 3.5a), corrected for multiple comparisons using cluster-based permutation. Nevertheless, correlations between the ERP difference and cueing benefit on either behavioural task at any of the different time points were not significant.

In the SWS group, ERP responses to the experimental cues showed a slow-oscillatory pattern that is characteristic of slow-wave sleep (see Figure 3.5b). Responses to the control cues showed this pattern to a lesser extent. However, this difference was not significant when controlling for multiple comparisons.



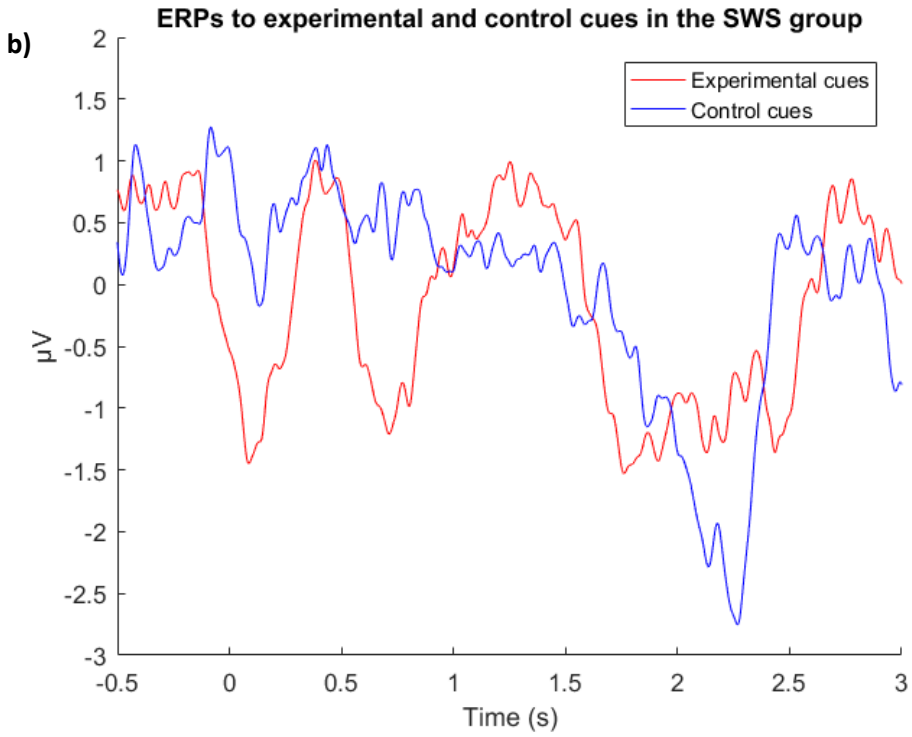


Figure 3.5. *Event-related potentials (ERPs).* ***a)*** *ERPs to experimental and control cues in the REM group. Grey rectangle indicates when the difference between these cues is significant (p < 0.05; corrected for multiple comparisons).* ***b)*** *ERPs to experimental and control cues in the SWS group.*

3.4.5 Time-frequency analysis

Time-frequency representation results in the REM group showed increased activity at approximately 13-17 Hz (roughly corresponding to the frequency of fast spindles) around 1.5 seconds after experimental cue onset. This activity seemed mostly absent in response to control cues (see Figure 3.6a and b). This difference was not significant when controlling for multiple comparisons.

Results in the SWS group revealed a strong response in the fast spindle-band (approximately 13-16 Hz) to experimental cues around 1.2 seconds after cue onset. The same response was much smaller for control cues (see Figure 3.6c-d), though this difference was a trend and not significant in a cluster-based permutation analysis ($p < 0.1$, see Figure 3.6e). This difference cluster did not correlate with behaviour, either on the learned or the remote associations (lowest $p = 0.263$).

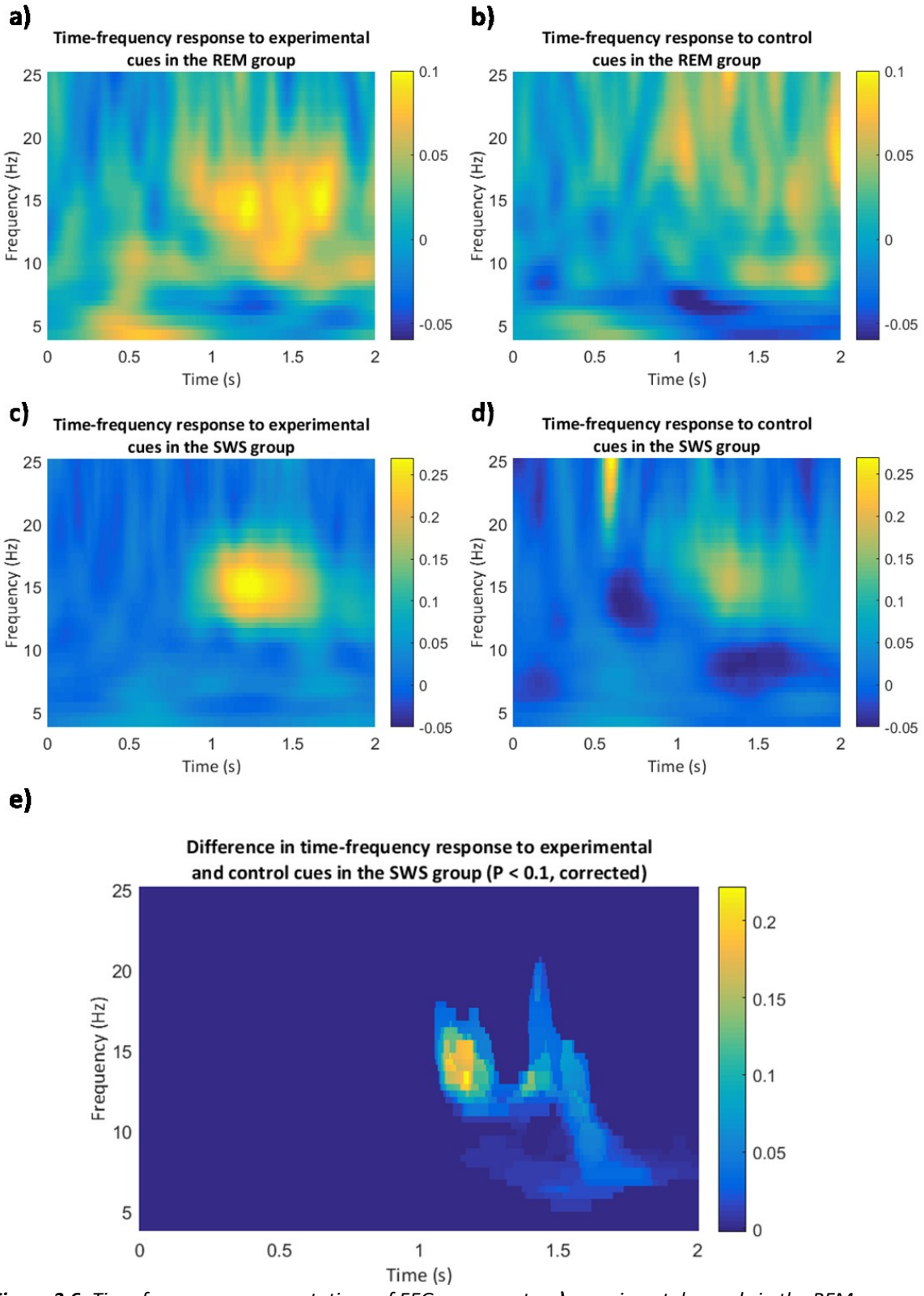


Figure 3.6. *Time-frequency representations of EEG response to* ***a)*** *experimental sounds in the REM group,* ***b)*** *control sounds in the REM group,* ***c)*** *experimental sounds in the SWS group,* ***d)*** *control sounds in the SWS group, and* ***e)*** *the difference between c) and d) at p < 0.1, corrected for multiple comparisons. Data represents an average of channels C3, Cz, C4, P3, Pz, and P4. Legends represent relative change from baseline.*

3.5 Experiment 1: Discussion

In this experiment, we investigated the effect of TMR of an associative inference task in both REM and SWS. Contrary to our expectations, learned associations were not strengthened by TMR during SWS. Moreover, we found no significant differences between TMR cues and control sounds in the SWS group, neither in the ERP nor the time-frequency results.

In contrast, we found that TMR during REM sleep increased performance on the remote associations. Specifically, this benefit for cued items appeared to increase over time, and was significant in a two-week follow-up. These results are in line with our predictions, although we had expected the difference between cued and uncued items to also be present the morning after participants had slept in the lab. It is possible that the TMR only led to a significant difference at the follow-up because forgetting over time enhanced the benefits that cueing provided. In this interpretation, the TMR during REM sleep served to protect those reactivated memories from forgetting. Another perhaps more exciting idea is that cueing during sleep set in motion a process that continued on subsequent nights, over time leading to a significant difference between cued and uncued items. This finding indicates that there is much to learn about the time course of TMR-related memory consolidation and associated benefits.

In terms of the electrophysiology, we predicted that we would find differences between our TMR cues and the control sounds without a strong memory component in our analyses of event-related potentials and time-frequency data, in both the REM and the SWS group. Our results showed no significant differences between these two types of sounds, except in the ERP results of the REM group. Here, we showed that ERPs between experimental and control sounds differed significantly in the REM group. This indicates that, during this sleep stage, the brain is able to distinguish task-relevant auditory cues from sounds that are not related to task performance, leading to distinctive processing of these two types of sounds. This result is discussed further in the general discussion at the end of this chapter.

These results represent some of the first direct evidence that memory reactivation during REM facilitates the formation of remote associations. In fact, besides this experiment, I know of only two other recent successful REM TMR studies: one showing that emotional arousal but not emotional memory is affected by TMR during REM sleep (Rihm & Rasch, 2015), and one which found that REM TMR enhanced both accurate and false recognition of faces (Sterpenich et al., 2014). In other words,

the discovery of a task that displays consistent REM-related improvements is important, and such a task would be very useful for examining the mechanisms of memory reactivation during REM sleep.

In addition, both the findings form the current study as well as those from Sterpenich and colleagues (2014) indicate that TMR during REM sleep may be important for the development of novel associations. These results provide initial evidence in favour of the BiOtA model, which proposes a role for memory reactivation during REM sleep in the detection of a shared underlying structure between different memory schemas (Lewis et al., 2018). Thus, our finding that remote associations are strengthened by REM TMR may represent a substantial step forward in our understanding of the role of this sleep stage in memory restructuring.

In other words, the results we obtained could have broad implications for the way we understand and study the role of REM sleep in memory. Therefore, we wanted to be confident that they are robust. The REM group in this experiment consisted of sixteen participants, which is a fairly common sample size for sleep TMR studies because these studies tend to show a large memory effect (e.g. Cousins, El-Deredy, Parkes, Hennies, & Lewis, 2014; Schreiner, Lehmann, & Rasch, 2015; Schreiner & Rasch, 2015; Simon, Gómez, & Nadel, 2018). Nevertheless, a post-hoc power analysis, using the effect size of our main REM result at the two-week follow-up, showed that we only obtained about 40% power with this sample size. Low statistical power not only reduces the chance that a study will detect a true effect, but also the probability that an obtained significant result reflects a true effect (Button et al., 2013; Fraley & Vazire, 2014). Thus, a replication of the REM group, with a larger number of new participants, would be more conclusive. Additionally, in light of the finding that results in psychology experiments often fail to replicate (e.g. Klein et al., 2014, 2018), we wanted to make sure that the results we obtained were replicable in an independent group of participants. Therefore, we set out to conduct a replication of the REM group.

3.6 Experiment 2: Materials and Methods

3.6.1 Participants

In this replication study, 27 participants were recruited. This sample size was based on a power calculation using the obtained effect size in the remote associations two-week follow-up. A power analysis was conducted using G*Power 3.1 (Faul et al., 2009) to test the difference between two dependent means, using a one-tailed test, an α of 0.05, and the obtained Hedges' g of 0.451. Results

showed that a total sample of 24 participants was required to achieve 70% power at this effect size. From our recruited sample, one participant was excluded for not reaching the required face-scene learning criterion (66%) before sleep, and another participant was excluded for having <5 hours of sleep during the night spent in the lab. One more participant was excluded for having <200 TMR cues played to them in the course of the night. The final sample, therefore, was 24 participants (13 females, mean age 22 ± 2.8 years). As the replication only concerned the REM effect, these participants were all allocated to the REM group. Participants were subject to the same inclusion criteria as the original study regarding lack of (sleep) disorders, medication, stimulant use, level of English, and behaviour before and during the study. This study was approved by the School of Psychology, Cardiff University Research Ethics Committee, and all participants gave written informed consent.

3.6.2 Experimental protocol

Participants underwent the same protocol as during the original study, including the two-week follow-up and the Cambridge Face Memory Test (Duchaine & Nakayama, 2006). In addition to the two-week follow-up, they were asked to come back for a second follow-up, which took place four weeks after the night spent in the lab. This follow-up was added to further explore the time course of TMR-related memory benefits. During this four-week follow-up participants performed both the inference (face-face) and recognition (face-scene) tasks again. All participants returned for the two-week follow-up, on average 13.5 days (range 11-19 days) after their monitored night. There was no difference in the follow-up time between this replication group and the original REM group ($t(38) = 0.04$; $p = 0.967$), as assessed with an independent samples t-test. All but one participant returned for the four-week follow-up, on average 26.7 days (range 21-32 days) after sleeping in the lab. During this replication, one small alteration was made in the inference (face-face) test. Originally, participants always saw a female probe, and had to pick the correct option from three male faces. In this replication, half of the participants saw a male probe with three female faces to choose from. This was then switched in the next sessions, such that a participant would see either a female probe (morning session) → male probe (two-week follow-up) → female probe (four-week follow-up), or male probe → female probe → male probe.

3.6.3 Targeted memory reactivation

Targeted memory reactivation took place in the same manner as the original study. Because all participants were part of the REM group, sounds were played exclusively during REM sleep. On

average 553.1 sounds were played during the night in this group of participants (each sound 13.8 times).

3.6.4 Polysomnography (PSG) data acquisition and analysis

Recordings were made with BrainVision Recorder and BrainAmp MR Plus amplifiers (Brain Products GmbH, Gilching, Germany) and sampled at 500 Hz. Twenty-one electrodes were placed on participants' scalp and face; at standard 10-20 system locations F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, P7, P8, O1, and O2. The remaining electrodes were two mastoid electrodes, left and right EOG, three EMG electrodes on the chin, and one ground electrode on the forehead. All electrodes were referenced to the mean of the two mastoid electrodes. Impedance was <5kΩ for scalp electrodes and <10kΩ for the EOG and EMG electrodes. PSG recordings were manually scored by two trained sleep scorers according to the standard AASM criteria (Berry et al., 2015). Both scorers were blind to the periods when sounds were replayed.

3.6.5 Data analysis

Behaviour in this replication experiment was analysed in the same manner as the original study. EEG, preprocessing, artifact rejection, time-frequency analysis, and ERP analysis were carried out using the same code as the original experiment. In this group, on average 9.7% of trials were removed during the artifact removal process.

3.7 Experiment 2: Results

3.7.1 Sleep data

Sleep data of this replication (REM) group can be found in Table 3.2. We conducted ANOVAs with between-subject factor group (combined with the groups from the original study) and the different sleep stages as dependent variables. Results show a main effect of group on time spent in Stage 2 ($F(3,68) = 9.25$; $p < 0.001$) and total sleep time ($F(3,68) = 8.38$; $p < 0.001$). It appears that this replication group spent less time asleep, and, likely consequently, less time in Stage 2.

Table 3.2. *Average minutes spent in sleep stages (± standard deviation) for the replication (REM) group.*

	Experiment 2 REM group
Stage 1	29.67 ± 23.55
Stage 2	192.85 ± 26.70
SWS	99.85 ± 28.02
REM	94.33 ± 27.02
Total sleep time	416.71 ± 42.95
Wake after sleep onset	28.31 ± 28.96

3.7.2 Learned associations (face-scene)

We conducted a repeated measures ANOVA with factors time (evening, morning, two-week follow-up, and four-week follow-up) and reactivation (cued and uncued). The dependent variable was accuracy in the face-scene test (i.e., learned associations). As in Experiment 1, results showed a main effect of time ($F(3,66) = 5.29$; $p = 0.002$, $\eta^2_G = 0.024$), which was caused by decreased performance at the two-week and four-week follow-ups (See Figure 3.7). Consistent with the findings of Experiment 1, there was no main effect of reactivation or any interaction with reactivation (all $p > 0.4$).

In addition, the benefit of TMR (i.e. the number of correct reactivated items minus the number of correct non-reactivated items) at any of the time points did not correlate with time spent in Stage 2, SWS, or REM, nor with the amount of TMR sounds played during the night (lowest $p = 0.209$, corrected).

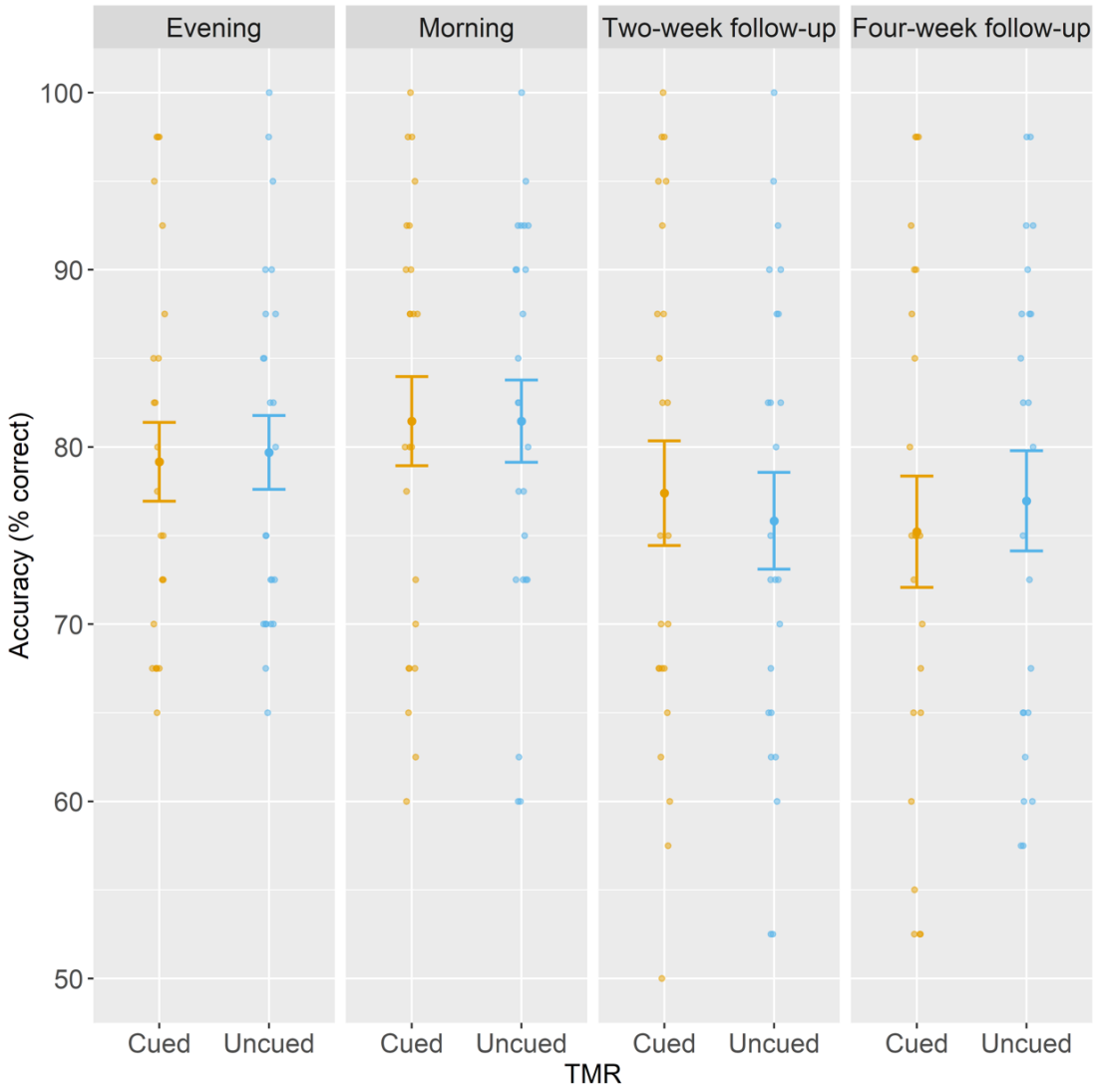


Figure 3.7. *Accuracy on the learned associations at different time points, in Experiment 2. Error bars represent 1 standard error of the mean (SEM).*

3.7.3 Remote associations (face-face)

Accuracy in the face-face test (i.e., remote associations) was assessed with a repeated measures ANOVA with within-participant factors time (morning, two-week follow-up, and four-week follow-up) and reactivation (cued and uncued). This analysis also showed a main effect of time ($F(2,44) = 11.36$; $p = 0.001$, $\eta^2_G = 0.066$). Participants showed decreased performance at the two-week follow-up as compared to the post-sleep measure, but accuracy increased again at the four-week follow-up (see Figure 3.8). In contrast to the findings of Experiment 1, there was no main effect of or any interaction with reactivation (all $p > 0.6$). Paired t-tests were conducted as planned comparisons on the difference between cued and uncued items. There was no significant difference at any time point (lowest $p = 0.417$, before correction).

Interestingly, the benefit of TMR (i.e. the number of correct reactivated items minus the number of correct non-reactivated items) at the two-week follow-up showed a trend towards a positive correlation with time spent in SWS (r = 0.54; p = 0.054, FDR corrected for multiple comparisons). However, closer inspection revealed that this correlation was caused by an outlier, and after removal of this participant the correlation did not approach significance (r = 0.33; p = 0.130, uncorrected). There were no other notable correlations with sleep stages, nor was there a correlation with the amount of TMR sounds played (lowest p = 0.138, corrected).

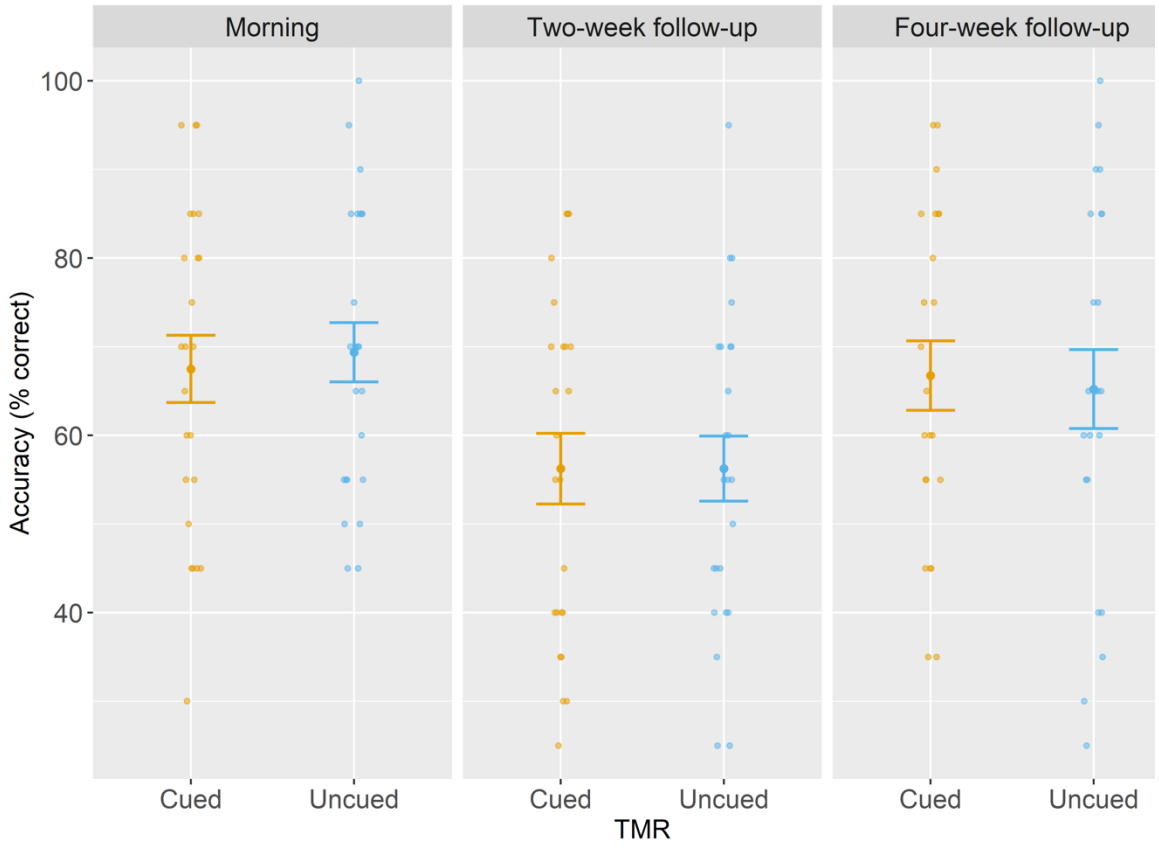


Figure 3.8. *Accuracy on the remote associations at different time points, in Experiment 2. Error bars represent 1 standard error of the mean (SEM).*

3.7.4 Event-related potential analysis

ERPs in response to experimental and control cues were also analysed for the replication group. The k-complex-like response found in the REM group of Experiment 1 was much smaller in this group and there was no difference between the responses to experimental and control cues (see Figure 3.9). The cluster-based permutation analysis did not reveal a significant difference.

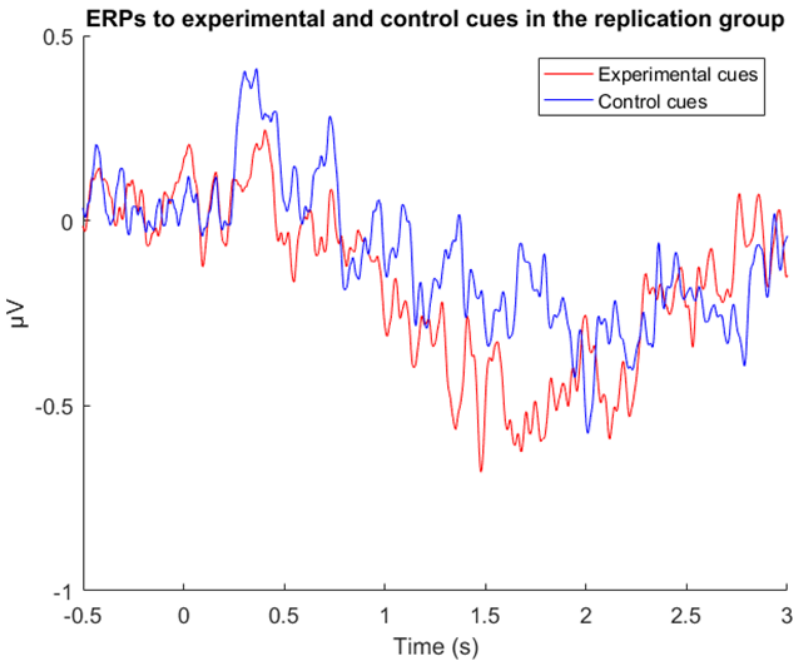


Figure 3.9. *Event-related potentials (ERPs) to experimental and control cues in Experiment 2.*

3.7.5 Time-frequency results

Time-frequency representation results in the replication group did not show the same fast spindle-band activity after experimental cue presentation as found in the previous REM group. Although somewhat similar activity did seem to be present (see Figure 3.10), it occurred earlier after the cue and was at a higher frequency. The difference in responses to the memory-related and control cues was not significant.

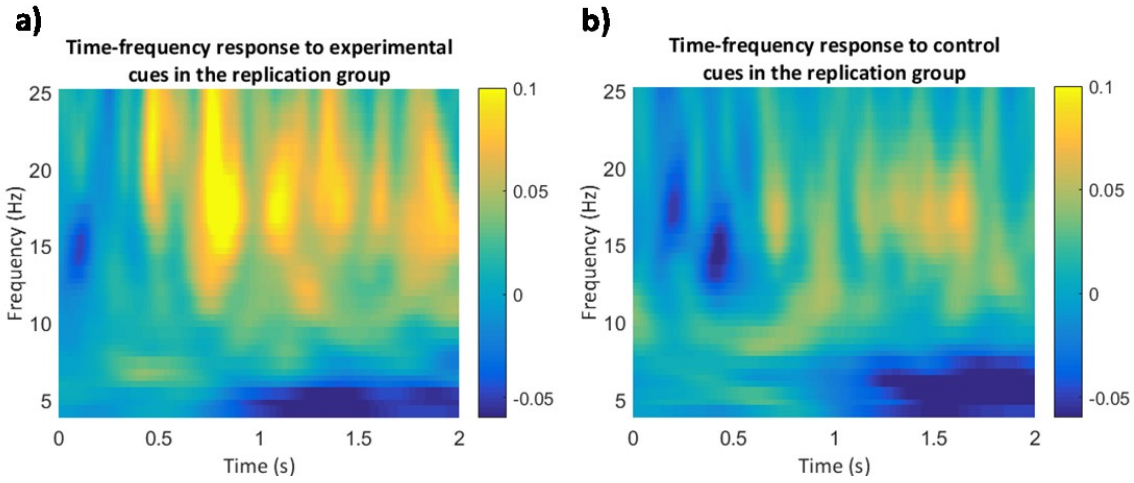


Figure 3.10. *Time-frequency representations of* ***a)*** *experimental sounds, and* ***b)*** *control sounds in Experiment 2. Data represents an average of channels C3, Cz, C4, P3, Pz, and P4. Legends represent relative change from baseline.*

3.8 Experiment 2: Discussion

In this replication, we found that TMR during REM did not affect performance on either the learned or remote association tests. Similarly, we found no difference between the ERPs of experimental and control sounds in this group of participants. Instead, we found a notable pattern in the results of our remote association test. Predictably, participants' performance was decreased at the two-week follow-up compared to their initial morning test, approximately to the same extent as in the original study. In contrast, in the four-week follow-up, their accuracy on the task returned to almost the same level as that of the test given the morning after they had slept in the lab. In contrast, performance on the learned (face-scene) associations was not similarly increased at the four-week follow-up.

We hypothesised that this pattern in the remote associations may have been caused by our manipulation of the gender of the probe face. If a participant was probed with a female face and had to choose the correct match from three male faces in their morning session, this was switched (to male and three female faces) at the two-week follow-up but then switched back at the four-week follow-up. Accordingly, participants' performance decreased when the genders were switched and increased when the genders were switched back, and thus appeared to be affected by this manipulation. If this was indeed the case, then this may have interacted with the TMR and consequently may have been at the root of the TMR null-results in this experiment. In other words, although Experiment 2 did not replicate the results of Experiment 1, we could not be certain about the cause of this disparity.

Therefore, to obtain clarity about the effect of REM TMR on remote associations and the replicability thereof, we conducted another replication of Experiment 1. In this instance, we first conducted an exact replication of the first experiment, meaning no face gender switching, up until the end of the inference task in the two-week follow-up. Subsequently, participants were asked to complete three more memory tests: a remote associations (face-face) test where the genders were switched, the original learned associations (face-scene) test, and a learned associations test where the probe was a face and participants had to choose the correct corresponding scene. This is described in more detail in the experimental protocol below. These extra tasks were added to

explore how participants' performance would be affected by such small manipulations in the way their memory was tested.

3.9 Experiment 3: Materials and Methods

3.9.1 Participants

In this second replication study, a total of 22 participants were recruited. One participant was excluded due to technical difficulties with the testing computer, and another participant was excluded because the TMR sounds had to be reduced to an almost inaudible volume in order to prevent persistent arousals. Thus, the final sample was 20 participants (11 females, mean age 21 ± 2.4 years). This replication again concerned the REM effect, which meant that all participants in this sample received TMR during REM sleep. Participants were subject to the same inclusion criteria as the original study regarding lack of (sleep) disorders, medication, stimulant use, level of English, and behaviour before and during the study. The study was approved by the School of Psychology, Cardiff University Research Ethics Committee, and all participants gave written informed consent.

Data collection was stopped after 20 participants (after exclusions) had been collected. We used a Bayesian analysis (computed with JASP, JASP Team, 2019) to evaluate the evidence for or against the null hypothesis given the sample size that was achieved. In contrast to null hypothesis significance testing, which leads to a binary decision about significance, a Bayesian approach allows for the calculation of relative evidence in favour of the null hypothesis or an alternative hypothesis (Wagenmakers et al., 2018). The results of this Bayesian analysis on our main data of interest (difference between cued and uncued item performance on the remote associations test at the two-week follow-up) gave a Bayes Factor (BF_{10}) of 0.192. A BF_{10} of below 1 indicates evidence against the alternative hypothesis – in this case, the hypothesis that performance on cued and uncued items is different. Our results were 1 ÷ 0.192 = 5.21 times more likely under the null hypothesis than the alternative hypothesis. This can be interpreted as moderate evidence in favour of the hypothesis that performance on cued and uncued items is *not* different (Schönbrodt, Wagenmakers, Zehetleitner, & Perugini, 2017). Using a Sequential Bayes Factor approach, it is also possible to visualise the accumulating evidence in favour of either hypothesis as each participant is added to the data (Schönbrodt et al., 2017). This showed that almost each additional participant in our sample moved the evidence towards support for the null hypothesis. In other words, it is unlikely (though not impossible) that the collection of additional participants would have led to evidence in favour of our alternative hypothesis.

3.9.2 Experimental protocol

Participants first underwent the same protocol as during the original study. However, after the inference (face-face) task was completed during the two-week follow-up, there was a slight deviation. Participants completed two extra tests, which were meant to evaluate the influence of the probe on memory performance in our tasks. Specifically, we wanted to examine whether switching the probe around would lead to memory gains such as those found in Experiment 2, and whether those gains would be affected by TMR. In the original inference task, participants saw a female face as the cue, and then had to pick the right male face out of three choices. In the extra inference task, the genders were switched around (male cue, three female choices). Additionally, in the original recognition test, participants were cued with the scene and sound they had learned, and then had to pick the right face out of four options. In the extra recognition test, the face was the cue, and participants had to choose which scene (out of four) belonged with that face. No sounds were played during this task (or during any of the inference tasks), as that would immediately give away the correct scene. The task order during the two-week follow-up was structured as follows: original inference → extra inference → original recognition → extra recognition. All but one participant returned for the two-week follow-up, on average 14.5 days (range 11-18 days) after the night spent in the lab. There was a significant difference in the follow-up time between this replication group and the original REM group ($t(33) = -2.62$; $p = 0.013$), as assessed with an independent samples t-test. Note, however, that the actual difference between the groups was only 1 day, as the average follow-up time in the original REM group was 13.6 days.

3.9.3 Targeted memory reactivation

Targeted memory reactivation took place in the same manner as the original study and the first replication. Because all participants were part of the REM group, sounds were played exclusively during REM sleep. On average 537.6 sounds were played during the night in this group of participants (each sound 13.4 times).

3.9.4 Polysomnography (PSG) data acquisition and analysis

EEG recordings were made with the same equipment and following the same procedures as in Experiment 2. PSG recordings were scored using the Z3 automatic sleep scoring algorithm (Patanaik, Ong, Gooley, Ancoli-Israel, & Chee, 2018) according to the standard AASM criteria (Berry

et al., 2015). This algorithm had an 84.23% agreement rate with the human sleep scoring in Experiment 1. The algorithm was blind to TMR sound onset.

3.9.5 Data analysis

Behaviour in this replication experiment was analysed in the same manner as the original study. To examine the overall REM effect on learned and remote associations, we also conducted combined analyses using all REM TMR groups. We further analysed reaction times (RTs) across experiments using ANOVAs and follow-up t-tests on the cued and uncued RTs as planned comparisons.

In terms of the EEG, preprocessing, artifact rejection, time-frequency analysis, and ERP analysis were carried out using the same code as the original experiment. In this group, on average 10.4% of trials were removed during the artifact removal process. To more directly investigate the connection between ERP results and behaviour, we also analysed the ERP of those participants which showed a behavioural effect of TMR. In addition, we separately analysed the behavioural data of those participants which showed the ERP effect we found in the REM group of Experiment 1.

3.10 Experiment 3: Results

3.10.1 Sleep data

Sleep data of this group can be found in Table 3.3. We conducted ANOVAs with between-subject factor group (combined with the groups from the original study) and the different sleep stages as dependent variables. Results show a main effect of group on time spent in Stage 1 ($F(3,60) = 6.01$; $p = 0.001$) and a non-significant effect on time spent in REM ($F(3,60) = 2.56$; $p = 0.063$). This replication group spent about 20 minutes less in Stage 1, and about 20 minutes more in REM sleep. This is likely a consequence of using a sleep scoring algorithm, as these occasionally have difficulty distinguishing Stage 1 and REM.

Table 3.3. *Average minutes spent in sleep stages (± standard deviation) for Experiment 3. Note that four participants were excluded from sleep stage (but not behavioural) analyses due to an issue with the EEG equipment near the end of the night, which resulted in that part of the night not being saved.*

	Experiment 3 REM group
Stage 1	16.41 ± 16.39
Stage 2	222.31 ± 41.71
SWS	90.88 ± 31.77
REM	121.50 ± 43.09
Total sleep time	451.10 ± 39.27
Wake after sleep onset	24.63 ± 23.40

3.10.2 Learned associations (face-scene)

As in the previous experiments, we conducted a repeated measures ANOVA with within-participant factors time (evening, morning, and two-week follow-up) and reactivation (cued and uncued). The dependent variable was accuracy in the face-scene test (i.e., learned associations). Results showed a main effect of time ($F(2,36) = 3.65$; $p = 0.036$; $\eta^2_G = 0.036$), which again seemed to be caused by decreased performance at the two-week follow-up (See Figure 3.11). There was no main effect of reactivation or any interaction with reactivation (all $p > 0.8$).

The benefit of TMR (i.e. the number of correct reactivated items minus the number of correct non-reactivated items) at any of the time points did not correlate with time spent in Stage 2, SWS, or REM, nor with the amount of TMR sounds played during the night (lowest $p = 0.138$, corrected).

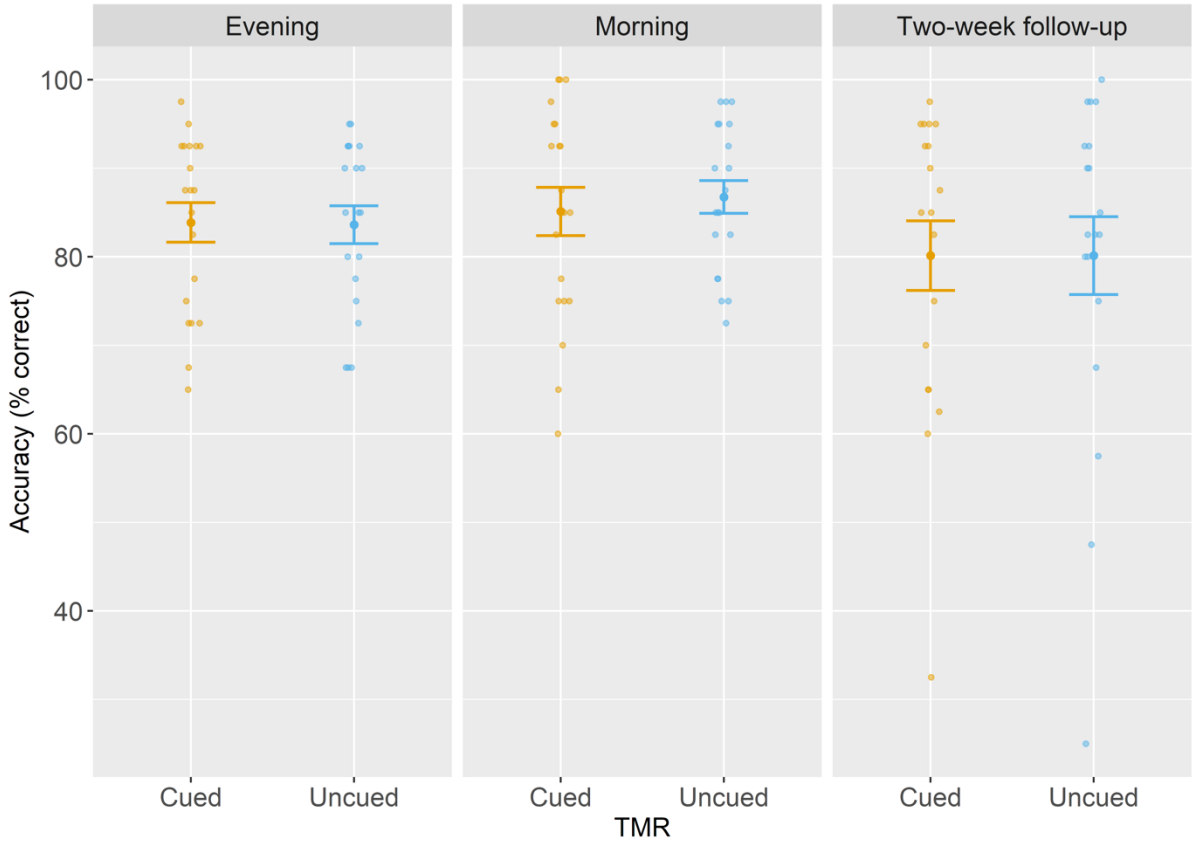


Figure 3.11. *Accuracy on the learned associations at different time points, in Experiment 3. Error bars represent 1 standard error of the mean (SEM).*

3.10.3 Remote associations (face-face)

To assess accuracy in the face-face test, we again conducted an ANOVA with factors time (morning and two-week follow-up) and reactivation (cued and uncued). Similar to the previous experiments, this analysis also showed a main effect of time ($F(1,18) = 5.23$; $p = 0.035$; $\eta^2_G = 0.045$) due to decreased performance at the two-week follow-up as compared to the post-sleep measure (see Figure 3.12). There was no main effect of or any interaction with reactivation (all $p > 0.2$). Paired t-tests were conducted as planned comparisons on the difference between cued and uncued items. There was no significant difference at any time point (lowest $p = 0.389$, before correction).

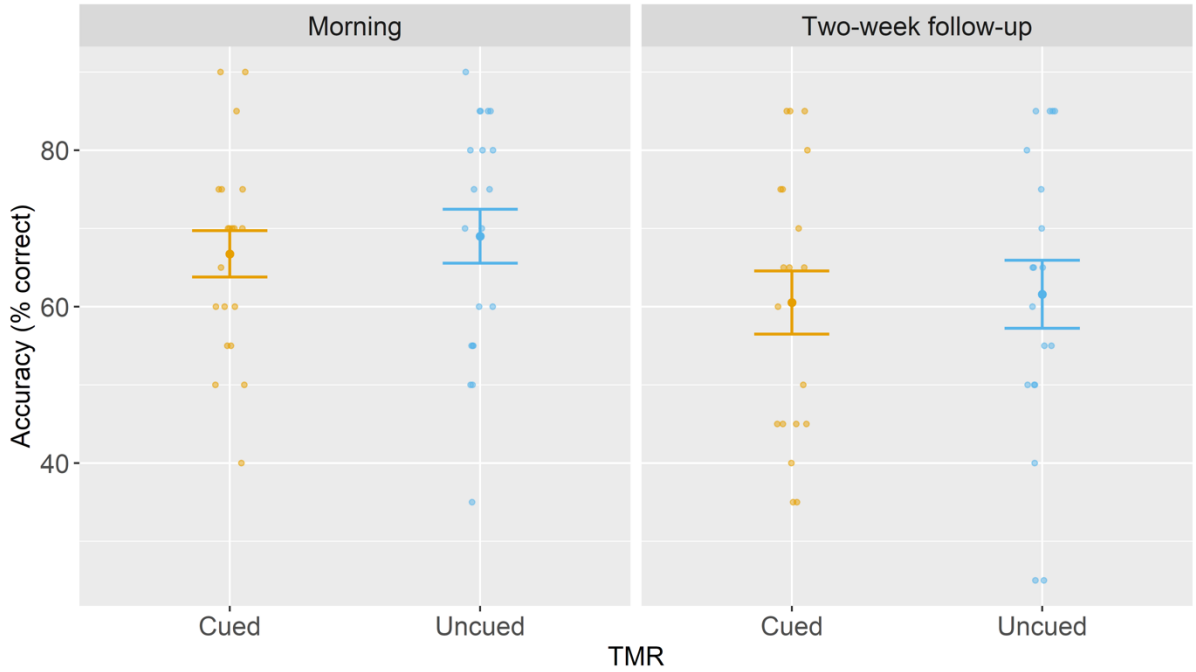


Figure 3.12. *Accuracy on the remote associations at different time points, in Experiment 3. Error bars represent 1 standard error of the mean (SEM).*

3.10.4 Extra behavioural analyses with probe switched

At the very end of the two-week follow-up, participants conducted an inference test with the gender of the faces switched (i.e. they now saw a male probe face and picked the correct matching face from a selection of three female faces), followed by a recognition test with the probe switched (i.e. they now saw a face and had to pick the correct corresponding scene). The results of these analyses are plotted in Figure 3.13. Note that one participant was removed from the switched recognition analyses, because they performed lower than chance (21.25%, where chance is 25%) on this task.

Participants showed very high accuracy on both of these tests. In the switched inference task this was approximately at the same level as performance on the normal inference task the morning after sleeping in the lab (accuracy of 68.03 ± 18.47 and 67.88 ± 14.27 percent for the switched and normal morning inference, respectively). This was also the case for the switched recognition task (accuracy of 87.99 ± 11.72 and 85.94 ± 10.31 percent for the switched and normal morning recognition, respectively). There was no difference between cued and uncued items in either task, as assessed with paired t-tests (lowest $p = 0.260$).

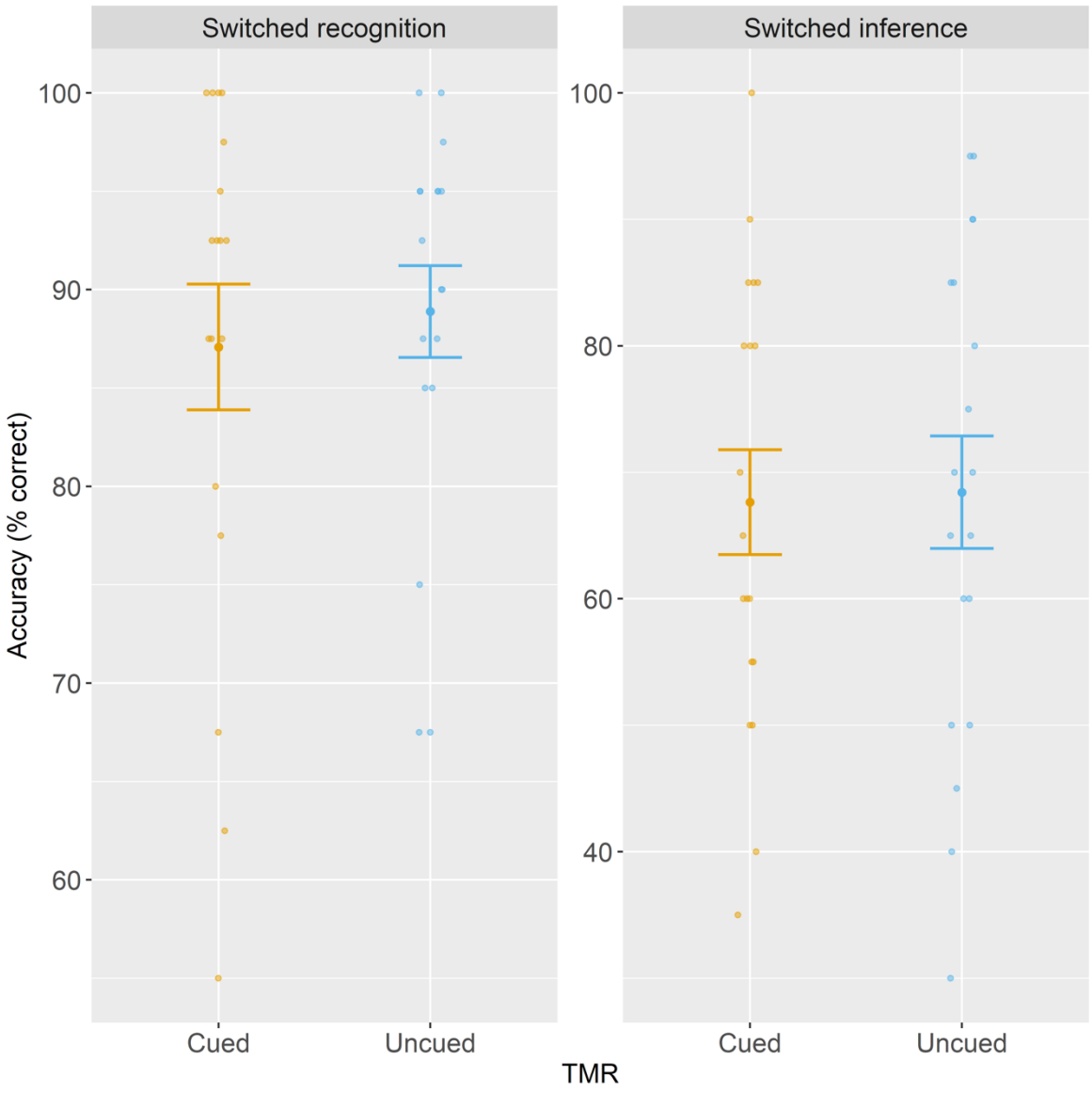


Figure 3.13. *Accuracy on the switched recognition and switched inference tasks, in the second replication group. Error bars represent 1 standard error of the mean (SEM).*

3.10.5 Event-related potential analysis

As in the previous experiments, ERPs in response to experimental and control cues were also analysed for the second replication group. The k-complex-like response found in the REM group of Experiment 1 appeared slightly earlier and much smaller in this group. There also appeared to be no difference between the responses to experimental and control cues (see Figure 3.14), and the cluster-based permutation analysis did not reveal a significant difference.

3.10.6 Time-frequency analysis

Time-frequency representation results in the second replication group shows quite a different pattern compared to the previous REM groups. In the previous groups the biggest increase in activity was seen after the experimental sounds. However, in this group there appears to be a relatively strong beta band response to the control sounds (see Figure 3.15). Nevertheless, the difference in responses to the memory-related and control cues was not significant.

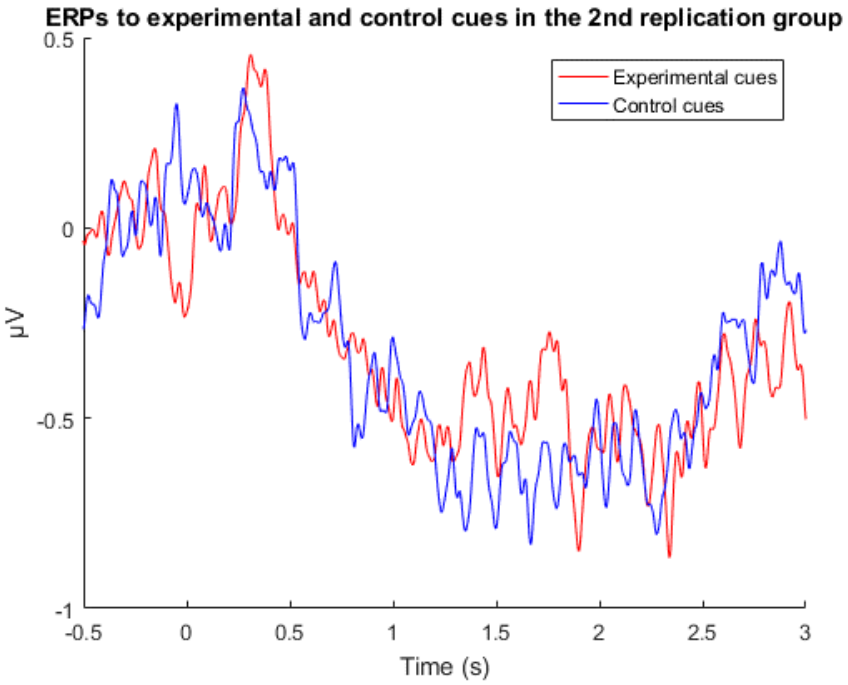


Figure 3.14. *Event-related potentials (ERPs) to experimental and control cues in Experiment 3.*

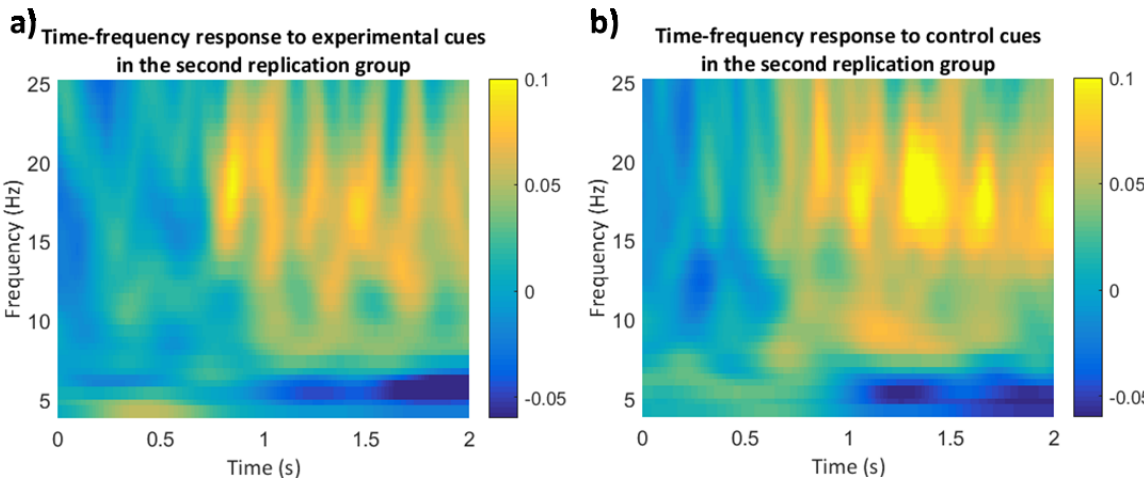


Figure 3.15. *Time-frequency representations of* ***a)*** *experimental sounds, and* ***b)*** *control sounds in Experiment 3. Data represents an average of channels C3, Cz, C4, P3, Pz, and P4. Legends represent relative change from baseline.*

3.10.7 Combined analyses

Combining the REM groups, learned and remote associations

We wanted to determine whether there was an overall effect of TMR during REM sleep. Therefore, we conducted several analyses where all groups that underwent TMR during REM sleep (i.e. the original REM group, and both replications) were pooled together. First, we re-analysed the learned (face-scene) associations using an ANCOVA with factors time (evening, morning, two-week follow-up), and TMR (cued and uncued). To control for any variation between the different REM groups, we added group (REM, replication 1, and replication 2) as a covariate. After adjusting for this group difference, there was a significant effect of time ($F(1.59,89.6) = 12.59$; $p < 0.001$; $\eta^2_G = 0.029$), but no significant effect of TMR or any interaction with TMR (lowest $p = 0.410$). When the different REM groups were added as a separate factor in the ANOVA, rather than as a covariate, this did not alter the findings. Thus, participants consistently showed forgetting over time, but there was no consistent effect of TMR in REM sleep on the learned associations.

We applied another ANCOVA to the remote (face-face) associations, using factors time (morning, two-week follow-up) and TMR (cued, uncued), with group as a covariate. This also revealed an effect of time ($F(1,56) = 32.85$; $p < 0.001$; $\eta^2_G = 0.061$), but no overall effect of TMR ($F(1,56) = 1.31$; $p = 0.258$). Furthermore, there was an interaction between the covariate of group and TMR ($F(2,56 = 4.70$; $p = 0.013$; $\eta^2_G = 0.012$), caused by the fact that the original REM group showed a TMR effect while neither of the replication groups did. We then omitted the covariate of group, to check the

overall effect of TMR without taking into account the differences between the groups. In line with the previous results, there was an effect of time ($F(1,58) = 36.01$; $p < 0.001$; $\eta 2G = 0.066$), but no overall effect of TMR ($F(1,58) = 0.52$; $p = 0.475$) and no interaction between time and TMR ($F(1,58) = 0.69$; $p = 0.410$). In short, similar to the learned associations, participants showed forgetting over time. However, there was no consistent effect of TMR in REM sleep on remote associations.

Bayesian analyses

We conducted a Bayesian analysis on the combined REM groups to determine the overall level of support for the null and alternative hypotheses with regards to our main result of interest: accuracy in the remote associations. These analyses were carried out using JASP (JASP Team, 2019). As mentioned before, the obtained Bayes Factor (BF_{10}) reflects the relative evidence for the alternative hypothesis over the null-hypothesis (Wagenmakers et al., 2018). This factor is a number which can vary from zero to infinity, where a BF_{10} of below 1 indicates evidence against the alternative hypothesis.

We first conducted a Bayesian repeated measures ANOVA with factors time (morning, two-week follow-up) and TMR (cued, uncued). The dependent variable was accuracy in the remote (face-face) associations test. This analysis resulted in a BF_{10} of 6.160e+7 for the effect of time, meaning very strong evidence in favour of an effect of time. In contrast, the BF_{10} regarding the effect of TMR was 0.182, meaning our results were $1 \div 0.182 = 5.49$ times more likely under the null hypothesis than the alternative hypothesis. This can be interpreted as moderate evidence in favour of the null hypothesis that there is no effect of TMR (Schönbrodt et al., 2017). Finally, the interaction between time and TMR resulted in a BF_{10} of 0.229, meaning our results were $1 \div 0.229 = 4.37$ times more likely under the null hypothesis than the alternative hypothesis. Thus, there was also moderate evidence against an interaction of time and TMR in our sample containing all REM participants.

Follow-up Bayesian paired t-tests comparing cued and uncued items in the morning and in the two-week follow-up were in line with the results of the Bayesian ANOVA. Results reflected moderate evidence against a difference between the cued and uncued items both in the morning ($BF_{10} = 0.142$) and in the two-week follow-up ($BF_{10} = 0.237$). Thus, the results of these Bayesian analyses overall indicate consistent moderate evidence against any effect of TMR on accuracy in the remote (face-face) associations test.

Reaction time analyses

Although we mainly expected changes in accuracy as a result of the TMR, we also explored reaction times because they may be reflective of a more implicit learning of the relationships between faces and scenes. For these analyses, we combined the groups that received TMR during REM, and conducted analyses on each TMR group (REM, SWS, Control) separately.

Looking at the learned associations (face-scene) in Figure 3.16, one can see that the reaction times decreased over time. For the REM group, we conducted an ANCOVA with factors time (evening, morning, two-week follow-up) and TMR (cued and uncued). We included experiment (original, replication, or second replication) as a covariate. Results show that there was a main effect of time ($F(1.37,76.93) = 53.91$; $p < 0.001$; $\eta^2_G = 0.199$). There was no effect of TMR or any interaction with TMR (lowest $p = 0.453$). We further conducted ANOVAs with factors time and TMR for the SWS and Control groups. Note that the factor time had three levels in the SWS group (evening, morning, two-week follow-up) and two levels in the Control group (evening, morning). In the SWS group, there was an effect of time ($F(2,30) = 10.08$; $p < 0.001$; $\eta^2_G = 0.122$), but no effect of or interaction with TMR (lowest $p = 0.436$). In the control group, there was a trend of time ($F(1,15) = 4.38$; $p = 0.054$; $\eta^2_G = 0.055$), but no effect of or interaction with TMR (lowest $p = 0.242$). Planned comparisons of cued and uncued items with t-tests showed no difference at any time point or in any group (lowest $p = 0.214$).

Figure 3.17 shows the remote associations (face-face). Again, it looks like participants generally became faster over time. The REM group was analysed with an ANCOVA with factors time (morning, two-week follow-up) and TMR (cued and uncued), and a covariate of experiment (original, replication, or second replication). This showed that there was indeed a significant effect of time ($F(1,56) = 13.30$; $p < 0.001$; $\eta^2_G = 0.054$). There was no effect of or interaction with TMR (lowest $p = 0.198$). In the SWS group, an ANOVA with factors time (morning, two-week follow-up) and TMR (cued, uncued) showed an effect of time ($F(1,15) = 7.06$; $p = 0.018$; $\eta^2_G = 0.105$), but no effect or interaction with TMR (lowest $p = 0.432$). The Control group only contained one time point on this task, which did not show an effect of TMR ($F(1,15) = 0.24$; $p = 0.629$). Planned comparisons of cued and uncued items with t-tests showed no difference at any time point or in any group (lowest $p = 0.323$).

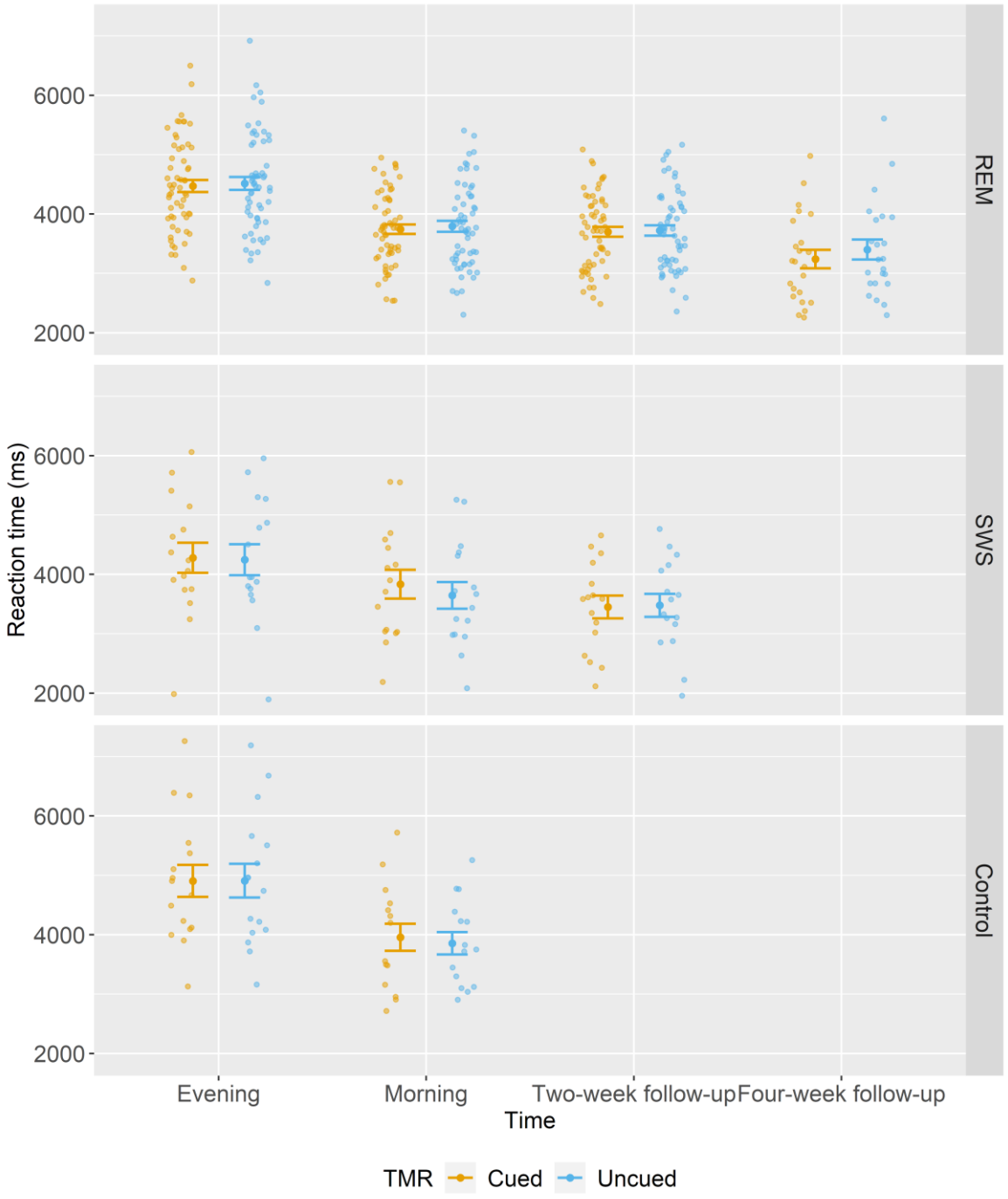


Figure 3.16. *Reaction time in milliseconds on the learned associations across different time points, separately for the REM, SWS, and Control groups. Note that for clarity, 4 data points above 9000 ms were removed from the plot. Error bars represent 1 standard error of the mean (SEM).*

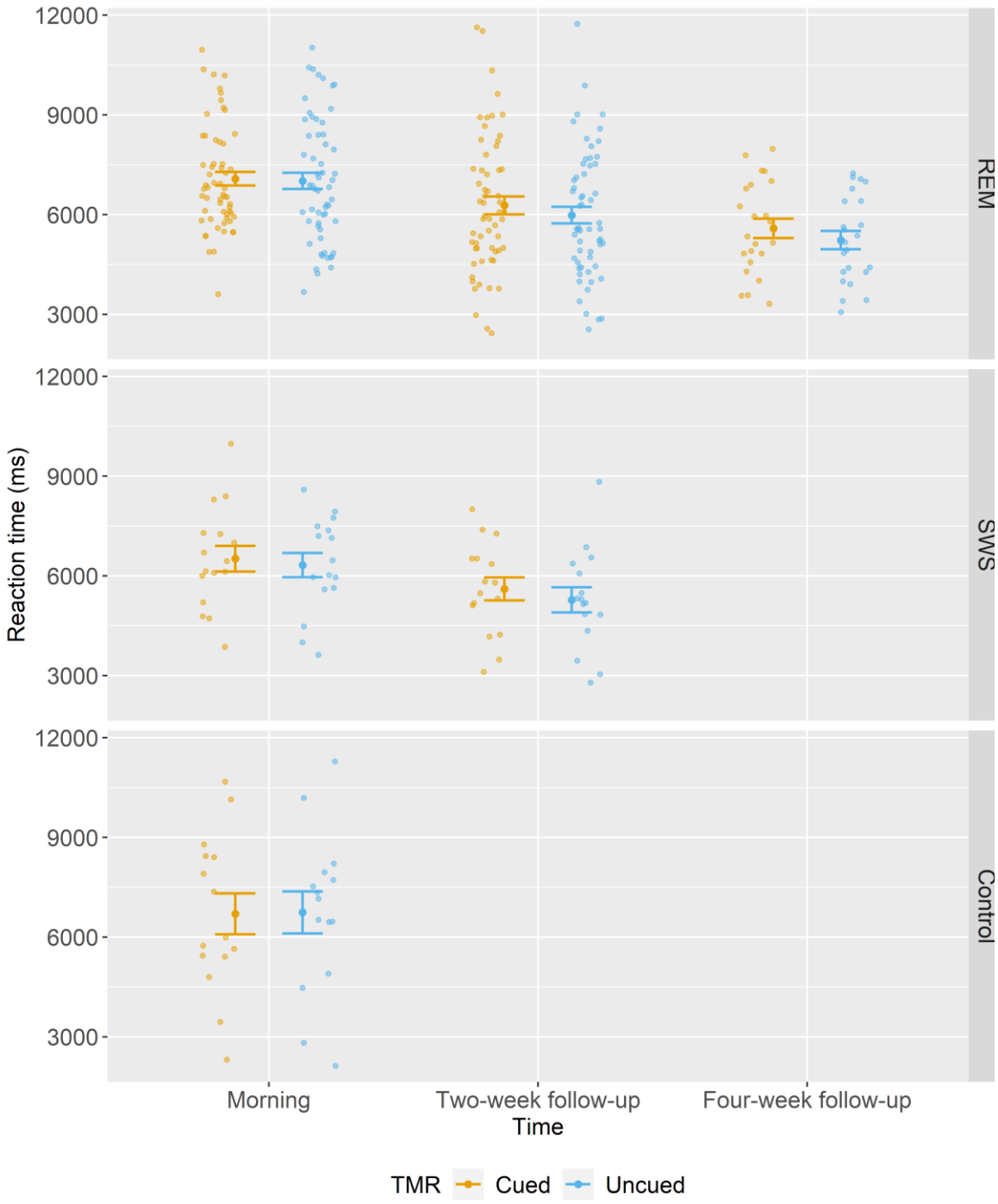


Figure 3.17. *Reaction time in milliseconds on the remote associations across different time points, separately for the REM, SWS, and Control groups. Note that for clarity, 5 data points above 12000 ms were removed from the plot. Error bars represent 1 standard error of the mean (SEM).*

Exploring the connection between ERPs and behaviour

The original REM group showed both a behavioural effect in the inference task and an ERP difference between experimental and control sounds. In the replication experiments, neither of these results were found, indicating that there may be a connection between the behavioural and electrophysiological results. It is possible that a difference in the way memory-related and control sounds are processed during the night reflects the extent to which experimental sounds are indeed processed, or even the degree to which a participant is susceptible to TMR. In other words, such electrophysiological processing may be associated with our behavioural results. If that is indeed the case, one would expect participants who showed a behavioural effect to show a bigger ERP difference during the night. Conversely, one would also expect that participants who show such an ERP difference between experimental and control sounds would show a bigger behavioural effect. To investigate this, we combined both replication groups and looked at each participant in detail. In this section, we thus explored the link between behavioural and electrophysiological results.

First, we examined whether those participants who showed a behavioural effect (that is, those participants that performed better on the cued compared to the uncued items in the two-week follow-up) would exhibit the ERP difference we found in the original REM group. Thus, we plotted and statistically analysed the ERPs of only those relevant participants in the replication groups. The result of this analysis is shown in Figure 3.18. Though the ERP difference appears slightly more pronounced in this combined group than in each of the replication groups separately, no significant difference was found between the ERPs of the memory-related and control sounds.

Conversely, we also examined the behavioural results of those participants in the replication groups that visually displayed a difference between the ERP of experimental and control cues. We conducted an ANOVA with factors time (evening, morning, two-week follow-up) and TMR (cued, uncued) on the learned associations, and one with factors time (morning, two-week follow-up) and TMR on the remote associations. Results predictably showed a main effect of time in the learned ($F(2,32) = 3.53$; $p = 0.041$; $\eta^2_G = 0.017$) and remote associations ($F(1,16) = 9.93$; $p = 0.006$; $\eta^2_G = 0.048$). There was no effect of TMR or an interaction (lowest $p = 0.225$).

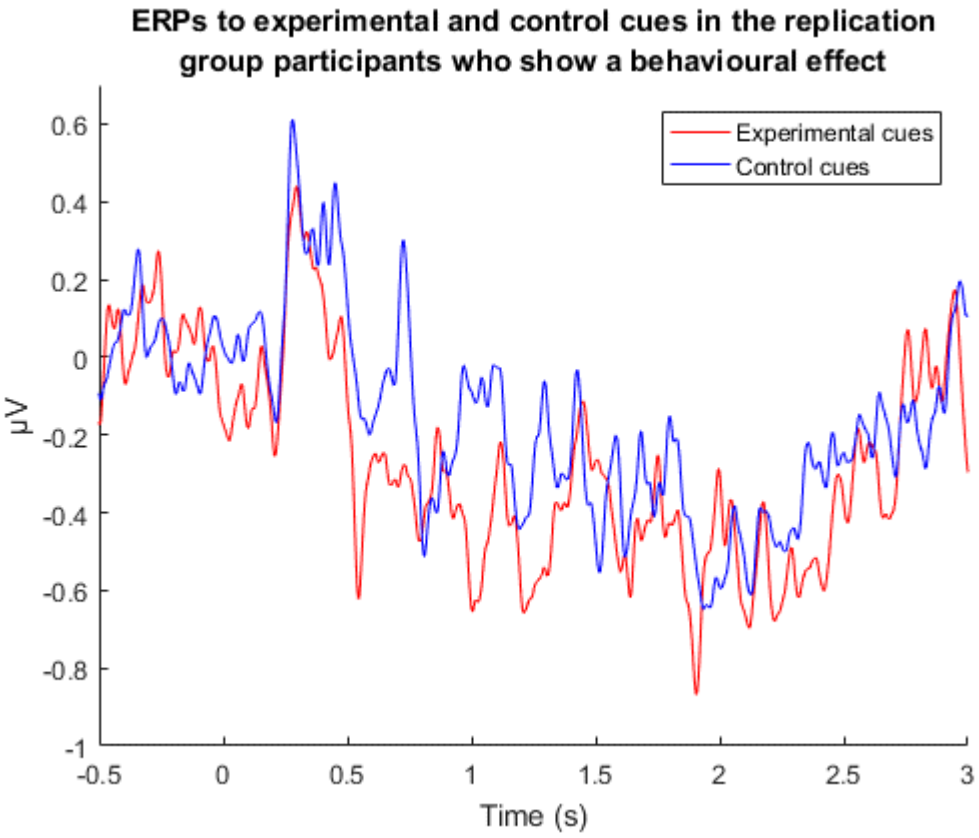


Figure 3.18. *Event-related potentials (ERPs) to experimental and control cues in a combined group of participants, consisting of those that showed a behavioural effect in the inference task at the two-week follow-up.*

3.11 Discussion

This study consisted of three experiments which were designed to investigate the effect of TMR of an associative memory task during sleep. In Experiment 1, we found that remote associations were strengthened by TMR during REM sleep in a two-week follow-up. Moreover, there was a difference in the ERP response to task-relevant and control sounds during this sleep stage. Because the results of this experiment were both exciting and important, but the sample size was relatively small, we decided to run a replication to increase confidence about the findings. In this second experiment, we did not replicate the original behavioural or ERP findings. Because the results of Experiment 2 were unexpected, and because we suspected that slight changes in the design might explain the differences between the two experiments, we conducted a second, more exact replication of Experiment 1. The results of this final experiment also did not replicate the original behavioural or ERP findings. Combined analyses and Bayesian analyses on all REM TMR subjects merged together indicated no significant effects of TMR and moderate evidence against any effects of TMR, respectively.

3.11.1 TMR during SWS does not improve learned associations

In the first experiment, contrary to hypotheses, no effect of TMR on behaviour in the face-scene task (learned associations) was found for any group. It had been hypothesised, based on the *BiOtA* framework (Lewis et al., 2018) and previous TMR experiments (e.g. Cairney, Guttesen, El Marj, & Staresina, 2018; Creery, Oudiette, Antony, & Paller, 2015; Rasch et al., 2007; Schreiner & Rasch, 2015; Shanahan, Gjorgieva, Paller, Kahnt, & Gottfried, 2018; Wang et al., 2019), that memory performance on the learned associations would be enhanced in the group that had received TMR during SWS. Our null-results therefore do not support the *BiOtA* model predictions.

It is possible that the specific structure of this experiment may have precluded us from finding strong enough results regarding these learned associations. Participants were required to learn the face-scene combinations to criterion before sleep: they had to have a 66% success rate when tested on these combinations. This was assessed by a multiple choice test after their first learning round. However, when they did not pass this test, a full second learning round followed, after which they would complete another test. Consequently, most participants far exceeded the learning criterion after this second test. Figure 3.3 shows that pre-sleep performance was on average 80%. This means that there were not many additional face-scene combinations participants could remember on top of their already high performance, and because they had learned them well they were also unlikely to forget many combinations. In other words, it may have been hard for the TMR to affect participants' high performance, either by increasing the amount of associations they remembered or by protection against forgetting. In subsequent studies, it may be beneficial to lower the learning criterion and make the experiment slightly easier (i.e. less face-scene combinations to learn). This way, most participants would only require one learning round before sleep. With one round of learning, participants may be more likely to forget items that are not strengthened with TMR. Furthermore, there would be more room for an increase in performance as well.

Another piece of the puzzle may be found in the electrophysiology. In the SWS group the time-frequency analysis showed an increase in fast spindles in response to experimental (memory-related) but not the control sounds, although this difference was not significant and did not correlate with TMR benefit. Nevertheless, spindles have frequently been implicated in memory improvements (e.g. Peyrache & Seibt, 2020; Ruch et al., 2012; Schabus et al., 2004; Zhang, Yetton, Whitehurst, Naji, & Mednick, 2020). Future studies could investigate whether TMR during Stage 2 has an effect on performance on this task, given that this sleep stage is particularly rich in sleep spindles.

3.11.2 TMR during REM sleep does not lead to consistent effects

Fitting with hypotheses and the *BiOtA* model described by Lewis, Knoblich, and Poe (2018), the first experiment showed that performance on the face-face task (remote associations) was improved after TMR during REM sleep. Interestingly, this effect seemed to increase over time, reaching significance in the follow-up that was conducted two weeks after participants spent the night in the lab. Unfortunately, this behavioural improvement was not found in either replication experiment. One could wonder whether a difference in the follow-up times between the groups could be at the root of this lack of replication. However, there was no difference in these times between the original REM group and the first replication group. There was a significant difference between the original REM group and the second replication group, but the actual difference in time was only 1 day (13.6 versus 14.5 days in the original and second replication group, respectively). It seems unlikely that one additional day could lead to a complete elimination of the TMR effect.

In the original group, data also showed an enlarged ERP to the control sounds as compared to the experimental sounds. This may reflect a kind of 'surprise' to hearing these sounds. Care was taken to expose participants to the control sounds during the course of the scene-sound learning sessions, to eliminate a startle response to hearing these sounds that might wake people up. Nevertheless, these sounds were only heard twice, whereas the memory-related sounds were heard at least ten times. Thus, the control sounds may have still been more surprising to hear during the night, which would reflect the difference in ERP response found. Although some ERP studies during sleep suggest a larger negative deflection to deviant stimuli (Brualla, Romero, Serrano, & Valdizán, 1998; Nordby, Hugdahl, Stickgold, Bronnick, & Hobson, 1996), others do note a larger positivity to surprising stimuli compared to frequent ones (Niiyama et al., 1994; Pratt, Berlad, & Lavie, 1999; Sallinen et al., 1996).

The difference in number of pre-sleep presentations (twice versus 10 times) could be considered a confound, and future studies looking to make similar comparisons would likely be better served by balancing these numbers. This issue does raise interesting questions about the nature of 'surprise' after sound presentations during sleep. Given that the experimental and control sounds were presented an equal number of times during sleep but not wake, the ERP difference during sleep could only have arisen due to the difference during wake. This could be a difference in memory content for both sets of sounds, or the difference in number of presentations. However, even if the ERP difference arose because of unequal presentation numbers, this implies that the brain remembered that fact during REM sleep. Thus, one could wonder whether the fact that the brain

recognises the control sounds as less common could be considered a memory in itself. Nevertheless, further research is required to back up such speculations.

Remarkably, this same ERP difference was not found in the electrophysiology of the replication groups, and indeed the time-frequency analyses show a different pattern as well. The reason for this electrophysiological difference between the groups is unclear, given the fact that sound tests and reports from participants practically exclude the possibility that they simply did not hear the sounds during the night.

3.11.3 Possible explanations for the disparity between the REM groups

Given the disparity between the different REM TMR groups, it is possible that there are vast inter-individual differences in susceptibility to TMR of this task. Even experiments that use the most robust tasks in terms of TMR effectiveness, for instance the vocabulary learning task or the serial reaction time task, contain participants that do not seem to benefit from TMR (Cousins et al., 2014, 2016; Schreiner & Rasch, 2015). Whether this is due to certain sleep characteristics, or because of some other reason, we do not know. Although there were no correlations with sleep stage durations in the current experiments, many studies have shown that there are large trait differences in these and other sleep characteristics (Buckelmüller, Landolt, Stassen, & Achermann, 2006; Purcell et al., 2017; van Dongen, Vitellaro, & Dinges, 2005; Werth, Achermann, Dijk, & Borbély, 1997). Other studies have found links between working memory and sleep-dependent memory consolidation (Fenn & Hambrick, 2012), and the amount of REM-NREM transitions a person goes through may also be involved (Kirov, Kolev, Verleger, & Yordanova, 2015).

It is possible that the original REM group included a large number of participants who were particularly responsive to TMR, whereas the replications did not. The fact that a significant difference between ERPs to memory-related and control sounds was only found in the original group also seems to suggest this. However, the link between the behavioural and ERP effects is still unclear. Even in the original experiment, correlations between behavioural results and the ERP difference were not significant. Moreover, when we looked at the ERPs of only those participants that showed the behavioural effect in the replications, no significant difference was found between the ERPs of control and experimental sounds. Similarly, the other way around, combining the behavioural data of those participants in the replication groups which show the ERP effect did not bring to light any behavioural effect of TMR.

Another factor that may interact with individual differences is task structure. Presented with the task as they were in the current experiment, it is possible that some participants adopted a competition rather than an integration approach to learning the faces. There exist several variations of the associative memory task used in this study. For instance, while we chose not to warn participants about or let them practice the inference (face-face) task that would take place in the morning, there are studies that do so (Preston et al., 2004; Zeithamova & Preston, 2010). It is possible that this would have promoted integration of the two faces, but our experiment was designed for implicit integration taking place during (REM) sleep rather than explicitly during encoding. There is also some evidence that strong learning of one face-scene pair before another is introduced promotes integration (Schlichting, Mumford, & Preston, 2015). On the other hand, one study has shown that increased integration during encoding actually eliminates the effect of sleep, potentially because the main function of memory consolidation during sleep may be the integration of new information into existing networks (Himmer, Müller, Gais, & Schönauer, 2017). Notably, it is unclear how these and other paradigm changes would interact with the TMR. Future experiments should investigate the effects of task design changes on sleep-dependent memory consolidation, as they may speak to the boundary conditions of the influence of sleep and TMR.

It is clear that small changes to the design of a task can have large effects on observed behaviour. For instance, in the first replication, we made a minor alteration that switched around the gender of the inference task probe across sessions. This apparently influenced participants to such an extent that we observed an increase in accuracy at the four-week compared to the two-week follow-up. In other words, participants performed better when the gender of the probe and test faces matched those of the initial inference task post-sleep. We reasoned that this alteration may have been the cause of the TMR null-results in Experiment 2. For instance, it could be reasoned that the gender switching allowed participants to see the face-face connections from both sides (female to male and male to female), forming a more complete circle of associations that may have overshadowed any TMR effects. Indeed, performance on our extra behavioural tasks in Experiment 3 (see section 3.10.4) was very high, illustrating that switching the probe could lead to a higher understanding of the associative relationships. However, in Experiment 3, a more exact replication of our original experiment, we again found no behavioural or ERP results of REM TMR. This indicates that task structure may have been less important than individual differences in this case.

It is important to consider the idea that there is simply no effect of TMR during REM sleep on remote associations. This would mean that the results of Experiment 1 in this chapter were a fluke finding,

potentially brought on by a lack of statistical power. As noted in section 3.5 this is a possibility, given the fact that low statistical power not only reduces the chance that a study will detect a true effect, but also the probability that an obtained significant result reflects a true effect (Button et al., 2013; Fraley & Vazire, 2014). This was one of the reasons that we conducted another experiment. Despite our best efforts, this experiment was likely also underpowered for any medium-sized or small effects. In other words, our most reliable results are those where we combine all subjects that underwent TMR during REM sleep. It has been noted that one would need at least 52 participants to show a properly-powered (80% power) statistically significant effect in a repeated measures t-test with two levels, at an effect size of d = 0.4, with more complex designs needing more participants (Brysbaert, 2019). With our total sample size of 60 REM participants, we should approach reasonable power to detect a main effect of TMR. As noted in section 3.10.7, the results of this ANOVA showed no effect of TMR, and indeed it is thus very possible that TMR during REM sleep has no behavioural effects in this task. Note, however, that to detect a null effect with 80% power in the same design as above, one would need a minimum of 215 participants (Brysbaert, 2019). It also remains curious that the first experiment contained two indications of TMR effects: a significant difference between cued and uncued items in the inference task in the two-week follow-up, and an indication of memory processing in the ERPs. This makes one hopeful that TMR during REM sleep could be effective under some conditions or in some participants.

3.11.4 The wider context of sleep effects on remote associations

This experiment was designed to investigate the role of sleep in promoting remote associations, which may be considered a first step in examining how sleep may be involved in creativity. Although some might question whether the inferences made in this task can be considered truly creative, it is clear that they require participants to move beyond memory retrieval into making a novel connection between items that have a shared underlying structure. As such, it is interesting to see how the current results fit in the broader sleep and creativity literature. In recent experiments using classical insight problems and magic tricks (Schönauer et al., 2018), riddles, anagrams, and visual change detection (Brodt et al., 2018), and an interactive computer game (Hołda et al., 2020), sleep was not shown to improve problem solving. However, although no correlations between sleep parameters and task performance were found, the sleep periods in these studies were naps, which generally contain little REM sleep. Moreover, none of these experiments used targeted memory reactivation.

On the other hand, there are also several studies that do demonstrate a benefit of sleep on creativity. For instance, Sio and colleagues used the remote associates task and found that their sleep group solved more difficult problems in a re-attempt than the other groups (Sio et al., 2013). Another study found that odour TMR leads to the generation and selection of more creative solutions to a problem presented before sleep (Ritter et al., 2012). Finally, Sanders and colleagues showed that auditory TMR during SWS boosted problem solving of cued versus uncued puzzles (Sanders et al., 2019). Note, however, that none of these experiments address the role of REM sleep specifically. The current study makes a first step to approach this, but many questions remain about the role of (REM) sleep in making remote associations and inferences, and promoting creativity.

3.11.5 Conclusion

While the REM group in our first experiment supports the *BiOtA* model idea that (REM) sleep promotes the finding of a shared underlying structure between different memories, the SWS group and the two replication REM groups do not. Whether this is because our task was not structured optimally, our participants were not TMR-sensitive, or because there is actually no effect, is as of yet unclear. However, given the lack of any effect in both replication groups, we must conclude that TMR during REM sleep does not reliably strengthen indirect associations in this task. This statement is substantiated further by our combined analyses, which indicated that TMR during REM sleep overall does not lead to significant behavioural effects. Bayesian analyses on the pooled REM subjects, finally, showed moderate evidence against an effect of REM TMR on remote associations. Thus, overall, the results of the experiments presented in this chapter do not support the *BiOtA* model.

CHAPTER 4

The effect of sleep and wakefulness on creativity

4.1 Abstract

Anecdotes and some studies have outlined the benefits of sleep for creativity. Nevertheless, it remains unclear which creative tasks benefit from sleep and what the mechanisms behind this are. Moreover, the effect of sleep on divergent thinking ability, i.e. the ability to come up with many different ideas, has not been adequately assessed. Participants took part in three sessions, each 12 hours apart, starting either in the morning or the evening. We thus investigated the within-participant effects of a period of sleep and a period of wakefulness on two creative tasks: the verb generation task (VGT) and the alternative uses task (AUT). As expected, semantic distance in the VGT was higher after an overnight interval than an over-day interval. However, overnight semantic distance change was not significantly higher than zero, while over-day change showed a trend towards being below zero. Surprisingly, performance on the AUT benefitted more from an over-day than an overnight interval, with over-day change being significantly higher than zero. Nevertheless, there was also an effect of time of day in this task. These results suggest that creative tasks that depend highly on semantic processing, like the VGT, may be particularly susceptible to sleep-related associative processes, possibly through an increase in spreading activation. Alternatively, this task may be sensitive to synaptic saturation which takes place during wakefulness, and experience a 'reset' of task performance after downscaling during sleep. On the other hand, more complicated tasks like the AUT, which are influenced by several strategies and processes, may instead benefit from an over-day interval which closer resembles a traditional incubation period, although more work is needed to distinguish the effect of time of day on this task.

4.2 Introduction

From the structure of benzene to an experiment that kicked off research into the chemical transmission of nerve impulses, many anecdotes assign a role for sleep in generating creative ideas (Mazzarello, 2000). When it comes to empirical evidence, there are some studies that have looked at the link between sleep and creativity. For example, one study showed that participants were more likely to gain insight into a task's hidden rule after a night of sleep compared to a night or a day of wakefulness (Wagner et al., 2004). Another group used targeted memory reactivation (TMR) with an odour to boost creativity (Ritter et al., 2012). Participants were given a problem that required a creative solution, and at the same time an odour was spread in the room. Those participants that were exposed to the same odour during problem presentation and sleep were more creative with their solutions, and better able to choose their most creative idea compared to

the other participants. In another TMR experiment, participants attempted to solve difficult puzzles before a night of sleep (Sanders et al., 2019). Each puzzle was associated with a different sound, and while participants were sleeping half of the sounds related to unsolved puzzles were played to them. In the morning, participants solved 31.7% of cued puzzles and only 20.5% of uncued puzzles, providing support for a role of sleep in problem solving.

Nevertheless, there are also studies that have found no effect of sleep on creativity. For instance, a study using classical insight problems and magic tricks found that sleep did not affect the general number of solutions, nor the amount of solutions reached by sudden insight (Schönauer et al., 2018). Other experiments have found that time spent awake is just as good for creativity as time spent asleep. One study looking at classical riddles, visual change detection, and anagrams gave participants two attempts to solve the problems (Brodt et al., 2018). These attempts were either right after one another, or after an incubation period of three hours that was spent asleep or awake. The incubation period increased solution rates in the classical riddles, and it did not make a difference whether this period was spent awake or asleep. Indeed, in the creativity literature there are many studies that indicate that incubation – taking a break from actively working on a problem – can increase creativity (for a meta-analysis, see Sio & Ormerod, 2009). Although incubation has traditionally been operationalised as a short break (i.e. a couple of minutes) spent awake (see e.g. the meta-analysis by Sio & Ormerod, 2009), the study by Brodt and colleagues (2018) shows that this period can also be several hours long and be spent asleep.

In summary, there is still substantial uncertainty about the effect of sleep on creativity. In the current study, we wanted to see whether a period of sleep and a period of wakefulness affected creative performance in two tasks. The experiment was set up in three different sessions, each 12 hours apart. Participants were allocated either to a group starting with a session in the morning, then the evening, and then the morning again, or to a group with sessions evening-morning-evening. Thus, each participant completed the tasks after an interval that contained wakefulness, and an interval that contained sleep.

The first task we used in this experiment was a verb generation task (VGT) which was adapted to be creative (Heinen & Johnson, 2018; Prabhakaran, Green, & Gray, 2014). Participants were given common nouns, and their objective was to come up with a verb in response. The verbs were generated in three different settings (common, random, and creative), which encouraged participants to explore their semantic network for these different noun-verb relationships. The VGT was chosen because of its purely semantic nature. Research on the consolidation of false memories

for words has suggested that these false memories are more likely to occur after sleep compared to wakefulness because sleep spreads activation from presented word representations to related concepts (Diekelmann et al., 2010; Monaghan et al., 2017; Newbury & Monaghan, 2019; Payne et al., 2009). This explanation has also been brought forward in two creativity studies, both looking at the remote associates test (Cai et al., 2009; Sio et al., 2013). Therefore, if sleep indeed spreads activation more widely than wake does, we would expect a period of sleep to lead to verb responses that are more distantly related to the stimulus noun. Performance on the VGT was assessed by looking at the semantic distance between the words. Quantitative measures of semantic distance have recently gained popularity in creativity research, because they are thought to allow a more direct and objective measure of the role of semantic memory in creativity (Kenett, 2019; Kenett & Faust, 2019; Prabhakaran et al., 2014). Thus, measuring semantic distance should be particularly useful when determining the spreading of participants' responses in a semantic network. In line with the idea of spreading activation, we expected semantic distance in the VGT to increase more after a night of sleep than a day of wakefulness in both the random and the creative settings. The common setting was included to explore whether this hypothesised effect of sleep would be similar when an increase in semantic distance would go against the objective of the task. We did not have specific predictions for this setting, besides the fact that overall semantic distance should be lowest when coming up with a common verb (compared to creative and random verbs).

With the exception of the study by Ritter and colleagues (2012), the sleep and creativity experiments mentioned above have predominantly looked at an aspect of creativity called convergent thinking. Generally, convergent thinking has been conceptualised as the process of generating one correct solution to a problem, such as in the remote associates test, puzzles, riddles, insight problems, and magic tricks (see e.g. Colzato, Ozturk, & Hommel, 2012; Gilhooly, Fioratou, Anthony, & Wynn, 2007). Divergent thinking, on the other hand, is considered to be a process that allows for the generation of many different ideas, usually in a context where multiple solutions can be correct (e.g. brainstorming). Divergent thinking has been particularly associated with real-life measures of creative achievement, such as the creation of plays and novels, and the attainment of patents (Plucker, 1999; Torrance, 1981). Moreover, the experiment by Ritter and colleagues (2012), which showed an effect of sleep with odour-induced task reactivation, consisted predominantly of divergent thinking (coming up with creative solutions to a problem). With this in mind, we were curious to see whether sleep would affect one of the most-used divergent thinking tasks: the alternative uses task (AUT).

Participants performed a computerised version of the AUT, wherein they were asked to come up with novel uses for a common household object. For example, a brick may be used to build something, as a doorstop, as a weapon, as a paper weight, etcetera. Although this appears to be a very simple task, participants were found to rely on many different mental functions to complete it, including episodic memory, semantic memory, and mental imagery (Gilhooly et al., 2007). The fact that several memory systems have been shown to affect this task makes it a good candidate for sleep-associated improvements, given the beneficial effect that sleep often has on memory (Rasch & Born, 2013). Thus, in line with the results obtained by Ritter et al. (2012), we expected participants' performance to increase more after a night of sleep than after a day of wakefulness.

4.3 Materials and Methods

4.3.1 Participants

Twenty-six native English speaking, non-smoking participants took part in this study (14 females, aged 18-35 years, mean age 22.8 ± 4.4 years). Participants reported no history of sleep, psychological, or neurological disorders, normal or corrected-to-normal vision and hearing, no use of any psychologically active medications, a lack of regular night work, and generally regular sleep. Because of a mental rotation task and a task involving analogies subjects were further required not to study (or work in the field of) mathematics, and have no more than three years of musical training in the last five years. Participants were asked to abstain from alcohol, caffeine, and napping from 24 yours before until the end of the experiment. Two participants were excluded from the study due to experimenter error regarding the task order and task counterbalancing, resulting in a final sample of 24 participants (12 females, 22.9 ± 4.6 years). This study was approved by the School of Psychology, Cardiff University Research Ethics Committee, and all participants gave written informed consent.

The sample size was based upon feasibility in relation to counterbalancing – to achieve good counterbalancing the study required 24 participants or a multiple of this number. We conducted a post-hoc power analysis to examine the achieved power in our result of interest: the comparison between overnight and over-day intervals in both the VGT and the AUT. We used G*Power 3.1 (Faul et al., 2009), a two-sided Wilcoxon signed-rank test (matched pairs), an α of 0.05, and the effect size dz calculated based on the means and standard deviations (SDs) of the groups (overnight vs over-day, respectively). In the VGT, the final sample of 23 participants had means of 0.012 and -0.014, SDs of 0.061 and 0.058, and a correlation of -0.022. The effect size was therefore 0.304. This

meant an achieved power of approximately 27.5%. In the AUT, the final sample size of 23 participants had means of 1.301 and 0.178, SDs of 1.749 and 1.631, a correlation of -0.10. The effect size was therefore 0.448. This meant an achieved power of approximately 51.7%. In other words, these experiments need to be replicated with a much higher number of participants to be able to make reliable statements about their results. To achieve 80% power, one would need a total sample size of approximately 91 and 43 participants for the VGT and the AUT, respectively.

4.3.2 Experimental Protocol

A schematic representation of the protocol can be found in Figure 4.1. The experiment consisted of three sessions, each 12 hours apart. Session 1 started either at 9am or 9pm, session 2 twelve hours after that, and session 3 twelve hours after session 2. Each participant was thus assigned to one of two groups: morning – evening – morning (M-E-M), or evening – morning – evening (E-M-E). It was ensured that the gender distribution in the two groups was approximately equal. Each session lasted approximately 1.5 hours and took place in a quiet room with a maximum of four participants at a time. Computers were separated by a divider screen.

At the start of each session, participants' alertness was assessed with the Karolinska Sleepiness Scale (KSS; Åkerstedt & Gillberg, 1990) and the Stanford Sleepiness Scale (SSS; Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973). Since mood may be related to creativity (Baas, De Dreu, & Nijstad, 2008; Davis, 2009), participants also completed a short Likert-style mood rating scale (based on Teasdale & Fogarty, 1979). Additionally, at the start of the first session, participants sleep quality and quantity over the past month was assessed with the Pittsburgh Sleep Quality Index (PSQI; Buysse, Reynolds, Monk, Berman, & Kupfer, 1989). During every session, participants completed five tasks: the Psychomotor Vigilance Test (PVT), the Verb Generation Task (VGT), the Alternative Uses Task (AUT), the Mental Rotation Task (MRT), and the Analogy Finding Task (AFT). Results and methods of the MRT and AFT are reported elsewhere. Verbal tasks (AUT, VGT, AFT) were always separated by non-verbal tasks (PVT, MRT) to minimise interference. Thus, tasks could be completed in twelve possible orders, which were counterbalanced with group (M-E-M or E-M-E). Task order was kept constant within-participant but pseudo-randomised (based on the above criteria) between-participant. All tasks and questionnaires were completed on computers with screen resolution 1920 x 1080, except for the AFT which was completed on paper. Audio during the AUT was presented with headphones (Sony MDR-ZX110NA). At the end of a participant's first evening session, they were given a portable dry-EEG headband (by Dreem, www.dreem.com) to wear while they were sleeping that night.

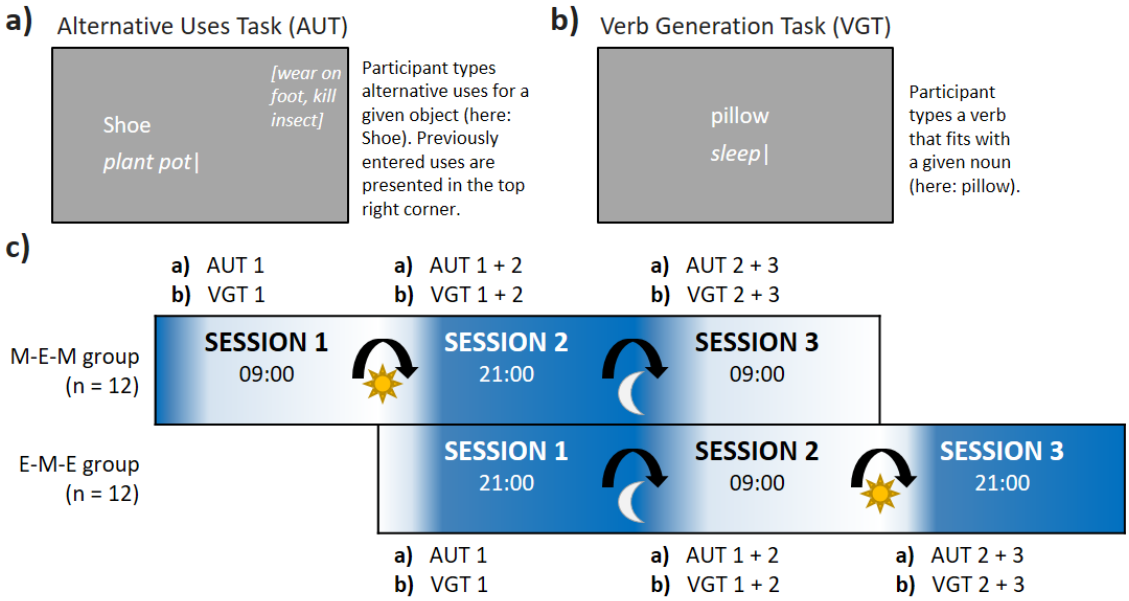


Figure 4.1. *Experimental procedures and tasks.* ***a)*** *A depiction of the computer monitor during the alternative uses task. Participants see the name of an object written on the screen, and their instructions are to come up with as many (and varied) uses for this object as they can think of within 3 minutes. After 3 minutes, the experiment moves on automatically.* ***b)*** *A depiction of the computer monitor during the verb generation task. Participants see a noun written on the screen, and their instructions are to come up with a verb that relates to the noun in one of three ways, depending on the part of the experiment (common, random, creative).* ***c)*** *Timeline of the experiment for the two groups in the experiment: Morning – Evening – Morning (M-E-M) and Evening – Morning – Evening (E-M-E). Depending on the group, participants started their participation in the morning or in the evening. They perform the AUT and VGT in each session, and, importantly, in sessions 2 and 3 they first re-complete the items/nouns that were introduced in the previous session.*

4.3.3 Verb Generation Task (VGT)

In the verb generation task (VGT), participants are presented with a common noun (e.g. *office*), to which they must respond with a semantically associated verb (e.g. *work*). A creative version of the task was developed by Prabhakaran and colleagues (2014). The variation of the task used in this experiment was based on a paper by Heinen and Johnson (2018). In this variation, participants carried out the VGT for three different measures. In the first, the *common* measure, participants were instructed to respond to the noun with a very common or typical verb. The second measure was the *random* measure, wherein participants had to come up with a completely unrelated verb response. Finally, in the *creative* measure, participants were told to generate a verb response that was creative – clearly related to the noun, but also rarely used in association with the noun. The complete set of instructions participants received can be found in Appendix A. These measures were designed to more closely examine the main facets of creativity: novelty and appropriateness. In the *common* measure, participants were expected to give responses high in appropriateness, but

low in novelty. On the other hand, the *random* measure was expected to yield highly novel but inappropriate responses. The *creative* measure, then, would be somewhere in the middle: novel responses that are still appropriate. The purpose here was to examine how each of these different measures would be affected by a period of sleep and a period of wakefulness.

The task was programmed in PsychoPy version 3.0.5 (Peirce et al., 2019) and stimuli were taken from Prabhakaran, Green, & Gray (2014). Care was taken to exclude any nouns that also occurred in the AUT or the AFT. Throughout the three sessions of the experiment 63 noun cues were used in total, a complete list of which can be found in Appendix B. Thus, in each session 21 new nouns were presented to participants, 7 in each measure (*common*, *random*, and *creative*). The nouns were randomised for each participant, based on one criterion: each collection of 7 nouns needed to contain a similar amount of low-constraint and high-constraint nouns. In other words, each 7 nouns (corresponding to one measure in one session) contained either three low-constraint nouns and four high-constraint nouns, or vice versa. Level of constraint was based on the frequency of the most common response. For instance, as Prabhakaran and colleagues (2014) explain, nearly all participants mention the verb *cut* in response to the noun *scissors*, which makes it a high-constraint noun. On the other hand, the most common response to the noun *tune* is the verb *play*, but this response occurs much less frequently in the overall sample, making it a low-constraint noun.

At the start of the task in each session, participants were shown general instructions for the task. They were told that the experiment would consist of three parts, and that in all of these parts they would see nouns and had to generate a verb to go with each of the nouns. It was further specified that they should only type one verb per noun, and that each part would also come with its own specific instructions. Then, the specific instructions for each respective measure were shown to participants before they had to complete that measure (see Appendix A). The measures were always completed in the same order (*common* → *random* → *creative*). This order was taken from Heinen & Johnson (2018), who argued that any noncreative condition that follows one where creativity is required will be influenced by demand characteristics. In each trial, participants were shown a noun, and they could type their verb response below it. When they pressed ENTER to move on, they were asked to type in all other verbs they thought of while generating their response. They were told to separate these by a comma, and press ENTER when they were done (or press ENTER if they had not thought of any other verbs while choosing their initial response). At the end of the trial, an instructions screen informed participants that they could now take a short break if

necessary. Moreover, it was noted that their next required response should again be only one verb. After pressing SPACE BAR to move on, the next trial was started.

In sessions 2 and 3, participants performed the same task again. In addition to 7 new nouns for each measure (new condition), they would also see the 7 nouns for that measure which they had newly seen the previous session (old condition). Thus, they responded to 7 x 3 = 21 nouns in the first session, and 7 x 2 x 3 = 42 nouns in sessions 2 and 3. For words they had seen before, they were told they could either type a new verb (if they felt it fit the instructions better), or type the verb they wrote before.

4.3.4 Alternative Uses Task (AUT)

In this study, a computerised version of the alternative uses task (AUT; Guilford, 1967) was used. The task was programmed in PsychoPy version 3.0.5 (Peirce et al., 2019) and involves coming up with alternative uses for a common household object that is written on the screen. In total, twelve household object were used, namely: ball, bottle, brick, button, fork, key, match, shoe, tin, towel, tyre, and umbrella. Over the three experimental sessions, each participant would come to see all of the objects, but the order was randomised for each participant. The task started with a short observation of half of the written objects paired with a sound clip of the spoken word. This was added as a sort of exploratory 'priming'. Participants were hereby exposed to several objects which would be used as stimuli in later sessions and could thus (consciously or subconsciously) start thinking of alternative uses for these objects. We were interested in whether scores for these items would subsequently be higher than those objects that were not presented at the start of the experiment. The objects seen and heard in this observation were chosen randomly for each participant. However, there were always six in total (half of all household objects used) and each of the three experimental sessions would contain two of these 'primed' objects, and two unseen ones. During the observation, each object was written on the screen for two seconds before moving on to the next object. The sound clips differed in length (due to the different word lengths), but were all between 300 and 450 milliseconds long.

After viewing and listening to six objects, the main task began. Participants were informed that this was a test of creativity, and instructed to list as many and varied uses that they could think of for the objects they would see. They were asked to focus on generating uses that were novel and appropriate. Finally, they were told that they would have three minutes per object to come up with and write down their answers. In each session, participants saw four new objects in randomised

order (new condition). Furthermore, in sessions two and three, they first saw the four objects that were new to them in the session before (old condition). For each of these, participants were asked to list uses for the objects as usual, but also to include whether this was an 'old' use that they had listed before (by writing /o after the use), or a new one they had just come up with (by writing /n). In other words, participants generated uses for four objects in session 1, and 8 objects each in sessions 2 and 3.

Participants pressed ENTER after each use they typed, and the uses previously generated for that object appeared in a list on the right upper corner of the screen. After three minutes, the experiment automatically moved on to the instruction screen again, and subsequently the next object when participants pressed the space bar.

4.3.5 Psychomotor Vigilance Test (PVT)

To assess fatigue-related changes in alertness across the three experimental sessions, we used the psychomotor vigilance test (PVT; Dinges & Powell, 1985). The PVT is a simple cued reaction time task. At the start of each trial, a white fixation cross was presented in the centre of a black screen. After a randomly chosen interval between 2 and 10 seconds, a rapidly upward counting timer was started. Participants were instructed to stop the timer as quickly as possible, using the space bar on the keyboard. When the timer was stopped, its number reflected the time (in ms) participants took to respond, and this was provided as feedback. If participants failed to respond within 2000 ms, participants were shown the text "Please pay attention" and the task moved on to the next trial. The task took 10 minutes in total and there was a break in the middle. The PVT was programmed in Matlab (R2015a, The Mathworks Inc., Natick, MA) with the Psychophysics Toolbox version 3 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007).

4.3.6 Data Analysis

Analyses were conducted and visualised using the R language and environment (version 3.6.3, R Core Team, 2020). To evaluate the effects of group, session, and measure on our outcome variables, we conducted mixed ANOVAs with the R package "afex" (Singmann et al., 2020). This package automatically applies the Greenhouse-Geisser correction for sphericity when Mauchly's test of sphericity is violated. Paired-sample t-tests were used as planned comparisons to determine the difference between overnight and over-day intervals. We used one-sample t-tests to evaluate whether overnight and over-day changes differed significantly from zero. Note that the Shapiro-

Wilk test revealed statistically significant deviations from normality in some of the sub-groups, though plotting showed that these deviations were small. Whenever this was the case, we used Wilcoxon signed-rank test rather than the parametric t-tests mentioned above. To assess effects of group on sleep and questionnaire scores, we used Welch's t-test, which does not assume equal variances and adjusts the degrees of freedom based on the size of each group and the variance within that group. It has been argued that this test should be used as a default rather than the Student's t-test when comparing independent groups (Delacre, Lakens, & Leys, 2017; Ruxton, 2006). Correlations were assessed using Pearson's correlation coefficient, or Spearman's Rho in the case of non-normal distributions. Statistical analyses were considered significant at $\alpha < 0.05$ and statistical tests were two-tailed. Corrections for multiple comparisons were done using the false discovery rate (FDR) method, which takes into account the expected proportion of falsely rejected hypotheses (Benjamini & Hochberg, 1995). Measures of effect size were included: generalised eta squared (η^2_G) for ANOVA as calculated with the "afex" R package (Bakeman, 2005; Lakens, 2013; Olejnik & Algina, 2003; Singmann et al., 2020), Hedges' g for t-tests as calculated with the "effsize" R package (Hedges, 1982; Lakens, 2013; Torchiano, 2020), and *r* for Wilcoxon tests as calculated with the "rcompanion" R package (Fritz et al., 2012; Mangiafico, 2020).

Sleep Analysis

Time spent in different sleep stages was assessed with an ambulatory dry-EEG device, i.e. the Dreem headband (provided and manufactured by Rythm). The device is made of a flexible band covered in fabric, which makes it adaptable to different head sizes and fairly comfortable. It uses five dry nanocarbon-coated fabric sensors at locations approximately corresponding to FPz, F7, F8, O1, and O2 to record EEG activity. The signal was recorded with a sampling frequency of 250 Hz and filtered with a band-pass filter of 0.4 – 18 Hz and two notch filters of 50 and 60 Hz, respectively. Data was post-processed for sleep stage detection using algorithms described in Debellemaniere et al. (2018). Accuracy of sleep stage classification of the automatic sleep scoring algorithm was comparable to that of trained sleep scorers: 83.5% for the algorithm versus 86.4% for the human sleep scorers (Arnal et al., 2020).

VGT Analysis

Participants' verb responses were inspected and any unambiguous spelling errors, suffixes, and extra words (e.g. 'to' in 'to walk') were removed. Further, there were several participants who appeared to have been confused about what a verb is, which meant that they occasionally wrote down adjectives or nouns instead. However, these participants' responses were overall not outlier and they were not excluded, with the exception of one participant who consistently displayed this confusion.[1] Thus, the sample size for this task was 23 participants.

Following previous research, we evaluated the responses using semantic distance (Heinen & Johnson, 2018; Prabhakaran et al., 2014). Like Heinen and Johnson (2018), we used latent semantic analysis (LSA; Landauer, Foltz, & Laham, 1998) to mathematically represent the distance between the nouns participants were given and the verbs they generated. Using the "LSAfun" package (Günther, Dudschig, & Kaup, 2015) in R (R Core Team, 2020), we calculated the average degree of co-occurrence between the two words in a large corpus of English language texts. We used the "EN_100k" corpus, which contains vectors for 100,000 words modelled on a corpus of approximately 2 billion words. It was downloaded from a repository on the website of the Universität Tübingen (www.lingexp.uni-tuebingen.de/z2/LSAspaces/). Semantic similarity between the words was calculated using the "Cosine" function, which computes the cosine of the angle created by the vectors for the noun and the verb. A word vector is created from its co-occurrence with other words in the corpus, and the vectors of words with similar meanings have smaller cosine distance. Semantic distance, then, was operationalised as the inverse of the semantic similarity value we calculated: 1 – semantic similarity (Prabhakaran et al., 2014).

After data curation and the LSA, 39 responses were not found in the LSA corpus (1.61% of the total responses). These were removed from the dataset. We calculated mean semantic distance scores per participant for each session, measure, and condition. We further calculated overnight and over-day change by subtracting performance on new items in the first session from old items in the second session, and new items in the second session from old items in the third session. Thus, change in the old condition reflected performance change on the same items, which could have been affected by offline memory processes. Performance in the new condition should be largely

[1] Note that the main VGT analyses were also conducted on a sample of 18 participants, excluding those that had on occasion been confused about what a verb is. These analyses led to the same conclusions in terms of significance or non-significance of the observed differences between overnight and over-day intervals, and in terms of significance or non-significance of deviations from zero.

free from such processes, and this could thus be used to check for the possible confounding factor of time of day.

AUT Analysis

Before scoring the items, we excluded one participant who had not followed the instructions (i.e., they wrote down free associations in response to the object, rather than uses). Thus, the sample size for this task was 23 participants. We further removed repeat responses, i.e. the same use typed twice.

Traditionally, the AUT is evaluated using three metrics: fluency, flexibility, and originality (Kaufman, Plucker, & Baer, 2008). Fluency was operationalised as the total amount of appropriate responses a participant provided. Flexibility was calculated by first allocating the responses to different categories. For instance, in response to the object *brick*, one could say *building a house*, *building a shed*, *weapon*, and *door stop*. This would be divided into three categories, as two of the responses belong to the same category. Participants received a point if they used a given category, and the flexibility score was the sum of those used categories. Originality, finally, was calculated based on the frequency of use of any given category. Participants received 1 point if they used a category that was used by <25% of participants, 2 points for a category used by <17.5% of participants, and 3 points if they used a category that was used by <10% of participants. All metrics were averaged over the four items a participant completed in each session, per condition (whether it was an old or a new item). We further calculated overnight and over-day change by subtracting performance on new items in the first session from old items in the second session, and new items in the second session from old items in the third session. Thus, change in the old condition reflected performance change on the same items, which could have been affected by offline memory processes. Performance in the new condition should be largely free from such processes, and this could thus be used to check for the possible confounding factor of time of day.

PVT Analysis

One participant had to be excluded from the PVT analyses, because their responses to the task were not correctly recorded by the script. Response times below 100 ms were considered a 'false start' and these were also excluded (Basner & Dinges, 2011). We did not exclude long response times, because they indicate low alertness and are thus meaningful.

4.4 Results

4.4.1 Sleep

Sleep data per group can be found in Table 4.1. All but one participant recorded data during the night. One further participant found the headband uncomfortable to sleep with and took it off halfway through the night. Finally, one participant slept less than 4 hours, and was excluded as an outlier on this basis. Thus, this table of sleep parameters contains data from twenty-one participants, eleven in the m-e-m group and ten in the e-m-e group. Welch's t-tests showed that time spent in sleep stages was similar for the different groups (all $p > 0.120$; right column).

Table 4.1. *Average minutes spent in sleep stages (± standard deviation), and p-values for the group difference.*

	Morning-evening-morning	Evening-morning-evening	Significance values of group difference
Wake	49.27 ± 26.79	48.90 ± 17.45	p = 0.970
Sleep onset	20.73 ± 11.10	18.10 ± 8.05	p = 0.540
Stage 1	0 ± 0	0.20 ± 0.63	p = 0.343
Stage 2	195.91 ± 39.29	196.10 ± 68.26	p = 0.994
SWS	124.64 ± 19.86	111.10 ± 18.23	p = 0.120
REM	118.91 ± 36.29	117.40 ± 37.55	p = 0.927
Total sleep time	439.91 ± 38.03	425.30 ± 60.54	p = 0.523

4.4.2 Questionnaires

All participants completed the PSQI, although two completed it during the second session rather than the first. However, given that the PSQI concerns sleep quality over the past month, these participants were not excluded from the analyses. The PSQI is scored on a scale of 1 to 21, with lower being better. In our sample, scores ranged between 1 and 7 points, with a mean of 4.46 (± 1.69), which indicates a good quality of sleep (Buysse et al., 1989). Participants in the m-e-m group had slightly better sleep on average than those in the e-m-e group (3.92 versus 5.0 points, respectively). Nevertheless, this difference was not significant ($t(21.27) = 1.62$; $p = 0.120$).

The rest of the questionnaires were completed by all but one participant, thus the sample size for these was 23 subjects. The KSS and SSS are scored on 9-point Likert scales, where 1 is most alert and 9 is least alert. Participants were comparably alert in the morning (score: 4.22) and the evening

(score: 4.43), as judged by the KSS. As expected, a paired t-test did not show a time-of-day effect ($t(22) = 0.46$; $p = 0.648$). Similarly, there was no notable difference between the morning and evening scores on the SSS (2.76 and 3.17, respectively), and a paired t-test did not show an effect of time-of-day ($t(22) = 1.79$; $p = 0.086$). Moreover, an unpaired Welch's t-test showed that KSS and SSS did not differ between the groups (lowest $p = 0.161$).

With respect to the mood rating scale, this consisted of three parts. Participants were asked to rate their current mood from happy (1) to unhappy (9), anxious (1) to calm (9), and despondent (1) to cheerful (9). Happiness was similar in the morning (score: 3.41) and the evening (score: 3.13). Participants were also calm in both the morning and evening (scores of 6.98 and 7.11, respectively). Finally, cheerfulness also did not differ between morning and evening (scores of 6.11 and 6.41, respectively). Paired t-tests were all non-significant for time-of-day effects (lowest $p = 0.304$). Similarly, the mood questions were not answered differently in the separate groups, as shown by unpaired Welch's t-tests (lowest $p = 0.108$).

4.4.3 Verb Generation Task (VGT)

Our analyses first focused on the old condition, where participants re-completed the task with nouns they had seen in the previous session. In this condition, offline memory processes thus had the opportunity to affect participants' responses to the nouns. We first compared the change in semantic distance over an interval containing sleep with that same change over an interval containing wake. Thus, we conducted an ANOVA with within-participant factors interval (wake or sleep) and measure (common, random, creative), and between-participant factor group (E-M-E or M-E-M). There was no significant effect of group ($F(1,21) = 0.02$; $p = 0.893$) or measure ($F(2,42 = 1.75$; $p = 0.185$). Importantly, however, the main effect of interval was significant ($F(1,21) = 8.13$; $p = 0.010$, $\eta^2_G = 0.049$), where an interval containing sleep led to significantly higher change in semantic distance than an interval containing wakefulness. This is illustrated in Figure 4.2 with a Wilcoxon signed-rank test ($V = 802$; $p = 0.015$, $r = 0.291$). There were no significant interactions (lowest $p = 0.300$). One-sample tests indicated that overnight change was not significantly greater than zero ($t(68) = 1.60$; $p = 0.114$), whereas over-day change showed a trend to being worse than zero ($V = 887$; $p = 0.056$).

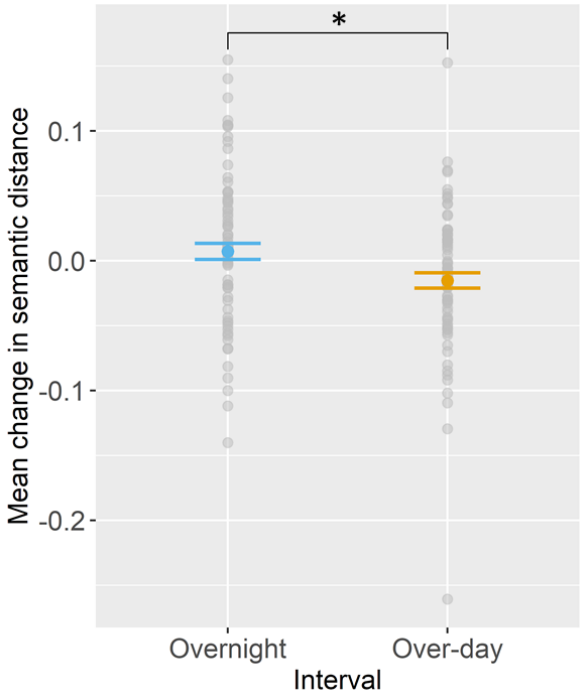


Figure 4.2. *Performance change overnight and over-day in the VGT, combined across outcome measures (common, random, and creative). * = V = 802; p = 0.015. Error bars represent 1 standard error of the mean (SEM).*

Although the effect of measure was not significant, we were interested in seeing whether there was a difference between sleep and wake intervals in each measure separately. Using paired t-tests, the difference between wake and sleep was only significant in the creative measure (t(22) = 2.39; p = 0.026; Hedges' g = 0.671), and this was reduced to a trend after multiple comparisons correction (t(22) = 2.39; p = 0.077; FDR corrected). These results are visualised in Figure 4.3. One-sample tests showed that overnight and over-day change did not significantly deviate from zero in any of the measures (lowest p, in the random measure over-day change: V = 75; p = 0.056).

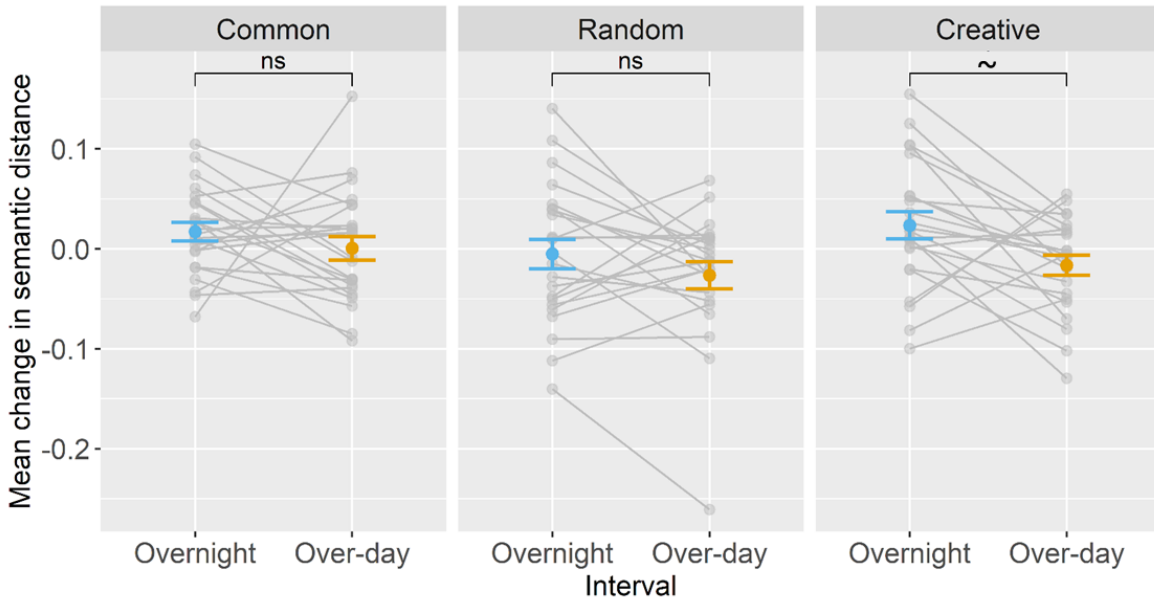


Figure 4.3. *Performance change overnight and over-day in the VGT, separated by outcome measure. ~ = t(22) = -2.39; p = 0.077; FDR corrected. Error bars represent 1 standard error of the mean (SEM).*

To make sure that the effects of sleep and wake intervals were not confounded by time of day effects, we turned our attention to performance in the new condition. This condition consisted of items that had not been seen before, and thus represents a more general creative ability which could vary by time of day. To investigate this, we conducted an ANOVA with between-participant factor group (E-M-E or M-E-M), and within-participant factors time of day (morning or evening) and measure (common, random, and creative). The dependent value was the semantic distance, where values for the morning and evening were averaged per participant. Crucially, there was no effect of time of day ($F(1,21) = 1.30$; $p = 0.267$) and no interaction with time of day (lowest $p = 0.203$). In fact, there were no significant effects or interactions (lowest $p = 0.203$), except for the effect of measure ($F(1.82,38.28) = 149.77$; $p < 0.001$, $\eta^2_G = 0.753$). In other words, although the measures were very different from each other, performance on all of them was approximately similar in the morning and in the evening.

The fact that we found a significant effect of outcome measure speaks to the validity of semantic distance as a measure of creativity. As seen in Figure 4.4, the highest semantic distance is reached when participants were instructed to think of a random verb, and the lowest when participants came up with a common verb in relation to the noun. In the creative measure, participants appeared to take into account the appropriateness component of creativity, and thereby their

semantic distances are reduced slightly compared to the random measure. These results are in line with those found in Heinen and Johnson (2018).

We further checked for an effect of session, with a separate analysis for each measure. We thus conducted three repeated measures ANOVAs, with within-subject factor session and between-subject factor group. There were no effects of session, group, or any interaction (lowest p = 0.223, after correction for multiple comparisons).

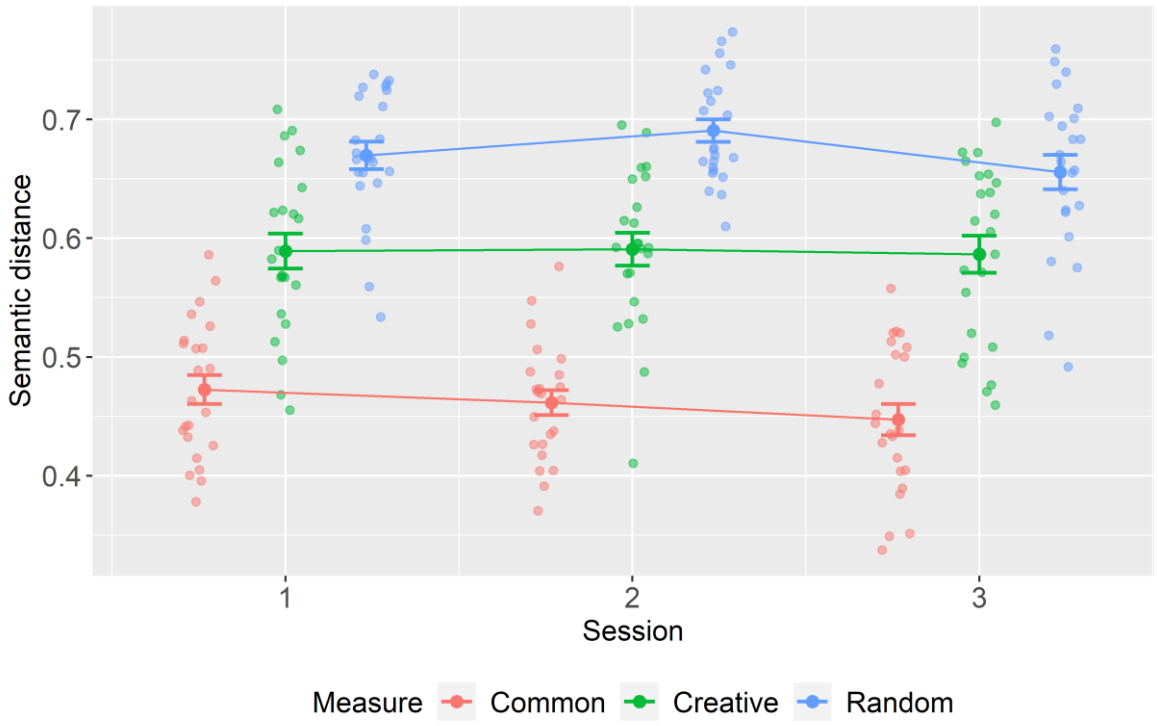


Figure 4.4. *Performance on the VGT per session, separated by measure (common, random, and creative). Error bars represent 1 standard error of the mean (SEM).*

To assess whether overnight change in performance was related to participants' sleep, we tested for correlations between this change in performance and time spent in Stage 2, SWS, and REM sleep. Performance on the random measure was positively correlated with time spent in Stage 2 (r = 0.562; p = 0.030, FDR corrected). No other correlations with sleep were significant (lowest p = 0.160, before correction).

Finally, we were interested in how changes in alertness related to our outcome measures. PVT performance was taken as a proxy for alertness. Thus, we also tested for correlations between

overnight and over-day change in performance on the VGT and change in PVT performance. None of these correlations were significant or approached significance (lowest p = 0.117, before correction).

4.4.4 Alternative Uses Task (AUT)

As in our analyses of the VGT, the main analyses of interest in the AUT focused on the old condition, where participants re-completed the task with objects they had already seen in the previous session. We first compared the change in fluency, flexibility, and originality over an interval containing sleep with change over an interval containing wakefulness. Thus, we conducted an ANOVA with within-participant factors interval (wake or sleep) and measure (fluency, flexibility, and originality), and between-participant factor group (E-M-E or M-E-M). There was no significant effect of group ($F(1,21) = 0.33$; $p = 0.574$) or measure ($F(1.57,32.97 = 0.71$; $p = 0.468$). Importantly, the main effect of interval was significant ($F(1,21) = 7.08$; $p = 0.015$, $\eta^2_G = 0.102$), where an interval containing wakefulness led to significantly higher change in score than an interval containing sleep. This is illustrated in Figure 4.5 with a Wilcoxon signed-rank test ($V = 1599.5$; $p < 0.001$, $r = 0.436$). There were no significant interactions (lowest $p = 0.131$). One-sample tests indicated that overnight change was not significantly different from zero ($V = 1394$; $p = 0.265$). On the other hand, over-day change was significantly greater than zero ($V = 1996$; $p < 0.001$).

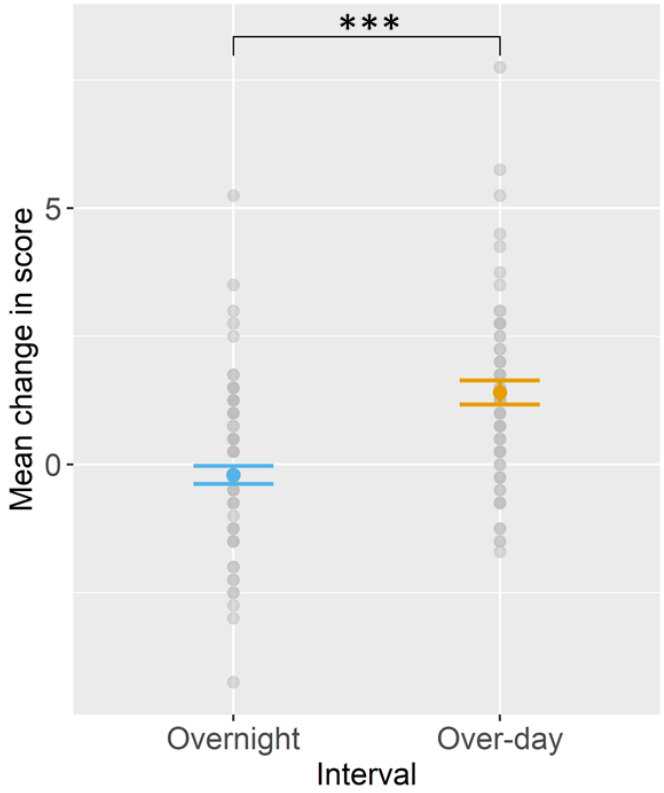


Figure 4.5. *Performance change overnight and over-day in the AUT, combined across outcome measures (fluency, flexibility, and originality). *** = p < 0.001. Error bars represent 1 standard error of the mean (SEM).*

Although the effect of measure was not significant, we were interested in seeing whether there was a difference between sleep and wake intervals in each measure separately. As Figure 4.6 shows, the difference between overnight and over-day change displayed a trend in fluency ($t(22) = 2.57$; $p = 0.052$; Hedges' $g = 0.818$; corrected) and flexibility ($V = 172.5$; $p = 0.074$; $r = 0.455$; corrected), but originality was not significant ($V = 176$; $p = 0.111$; $r = 0.309$; corrected). One-sample tests showed that overnight change was not significantly different from zero in any of the measures (lowest p, in the flexibility measure: $t(22) = 1.36$; $p = 0.189$). In contrast, over-day change was significantly higher than zero in all measures: fluency ($t(22) = 5.19$; $p < 0.001$), flexibility ($V = 250.5$; $p < 0.001$), and originality ($V = 180$; $p = 0.026$).

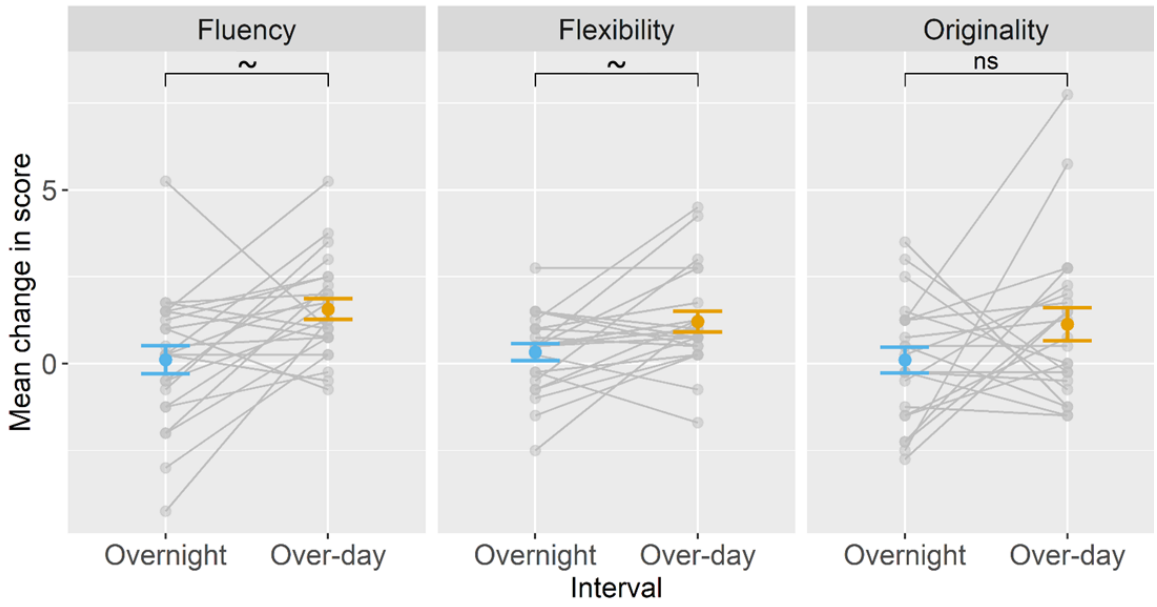


Figure 4.6. *Performance change overnight and over-day in the AUT, separated by outcome measure. ~ = p < 0.010, specifically p = 0.052 (Fluency) and p = 0.074 (Flexibility), after correction with the FDR method. Error bars represent 1 standard error of the mean (SEM).*

To check for time of day confounds, we again looked at performance on the new items. We conducted an ANOVA with between-participant factor group (E-M-E or M-E-M), and within-participant factors time of day (morning or evening) and measure (fluency, flexibility, and originality). The dependent value was the performance score, where values for the morning and evening were averaged per participant. As expected, there was a main effect of measure ($F(2,42) = 24.07$; $p < 0.001$, $\eta^2_G = 0.086$). This was expected, because fluency (the total amount of uses a participant comes up with) is almost always higher than flexibility (the amount of different categories in which a participant's uses fall into). Unexpectedly, there was also a main effect of time of day ($F(1,21) = 4.54$; $p = 0.045$, $\eta^2_G = 0.016$), with scores being higher in the evening than in the morning. There were no other main effects or interactions (lowest $p = 0.238$).

Unfortunately, as Miller and Chapman (2001) explain, there is statistically no manner in which we can control for this effect of time of day in our main analysis of interest, i.e. the analysis of the effect of a sleep or wake interval on AUT performance. Because overnight and over-day changes are inherently linked with time of day, there is no statistical way to "unconfound" them in our sample. Although a covariate analysis would remove some (shared) variance due to time of day, this would systematically distort the results relating to overnight and over-day changes. Thus, our finding that

an over-day interval leads to higher scores on the AUT than an overnight interval is likely partly attributable to time of day effects.

We wanted to see whether overnight change in performance was related to participants' sleep parameters. Thus, we correlated performance on the different measures with time spent in Stage 2, SWS, and REM. Overnight originality change was negatively related to time spent in Stage 3 sleep, though this was not significant (r_s = -0.409; p = 0.073), but this trend was absent after multiple comparisons correction (p = 0.219). There were no other correlations that approached significance (lowest p = 0.147, before correction).

We were further interested in how changes in alertness related to our outcome measures. PVT performance was taken as a proxy for alertness. Thus, we also tested for correlations between overnight and over-day change in performance on the AUT and change in PVT performance. Change in alertness was significantly negatively correlated with originality change (r_s = -0.416; p = 0.015, FDR corrected). In other words, when participants became more alert (i.e., faster on the PVT), they also scored higher on originality in the AUT.

Lastly, we investigated whether semantic distance values in the VGT were related to performance on the AUT. There were no significant correlations between any of the measures separately, nor when the measures were combined into average scores for each task (lowest p = 0.239).

4.4.5 Psychomotor Vigilance Test (PVT)

Because performance on the PVT, as a proxy for alertness, was related to change in the AUT originality, it was important to analyse PVT performance more closely. An ANOVA with between-subject factor group (M-E-M, E-M-E) and within-subject factor time of day (morning, evening) showed that there was no effect of time of day (F(1,21) = 0.17; p = 0.683). There was also no effect of group, nor any interaction between the factors (lowest p = 0.400). Figure 4.7a shows that reaction time was slightly higher across all sessions in the e-m-e group, and reaction time increased slightly across the sessions. However, an ANOVA with factors group (between-participant) and session (within-participant) indicated that there was no main effect of group (F(1,21) = 0.25; p = 0.621) or session (F(2,42) = 1.87; p = 0.167), nor an interaction between the two (F(2,42) = 0.12; p = 0.890).

Furthermore, as Figure 4.7b shows, overnight and over-day change in PVT reaction time was minimal, but this varied a lot between subjects. There was no significant difference in change between the two intervals, as shown by a Wilcoxon signed-rank test (V = 125.5; p = 0.715).

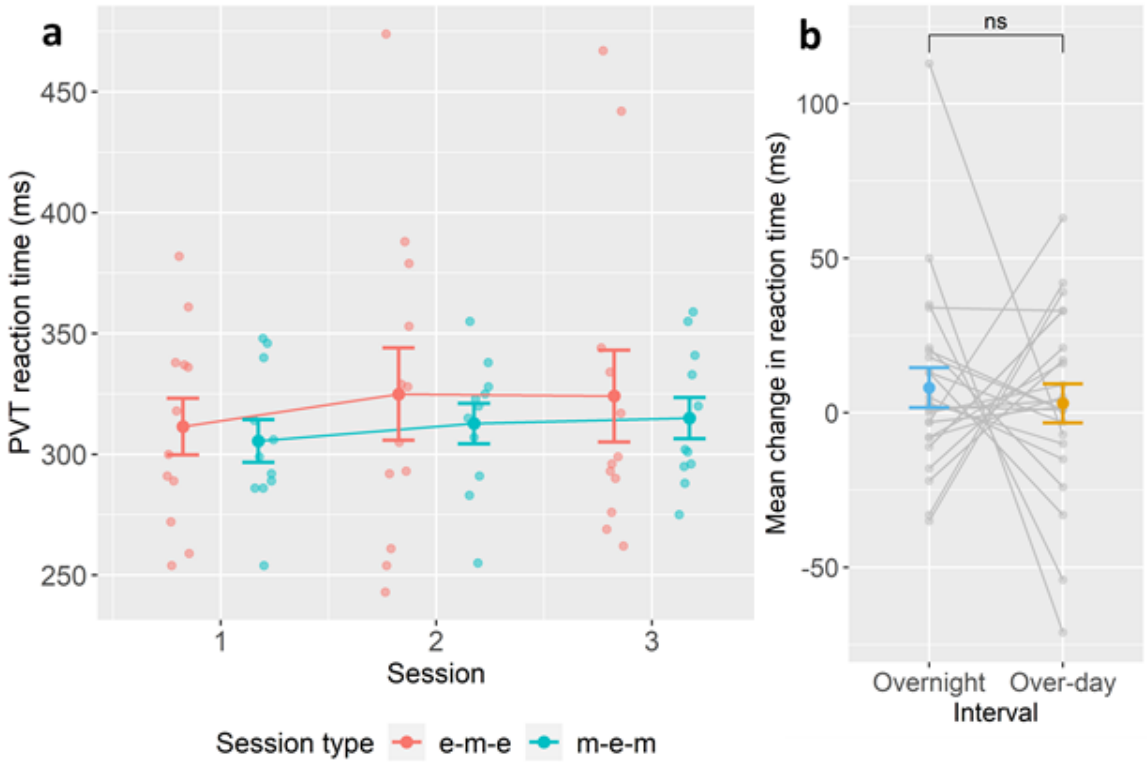


Figure 4.7. *Performance on the PVT* ***a)*** *per session, and* ***b)*** *separated into change overnight and over-day. ns = not significant, V = 125.5, p = 0.715. Error bars represent 1 standard error of the mean (SEM).*

4.5 Discussion

In this chapter, we set out to investigate whether a period of sleep and a period of wakefulness affected creative performance in two tasks. In the VGT, an overnight interval increased semantic distance more than an over-day interval. However, overnight semantic distance change was not significantly higher than zero, while over-day change showed a trend towards being below zero. When we separated the results by measure, the difference between a wake and sleep interval was only significant in the creative measure, and this was reduced to a trend after correction for multiple comparisons. Per-measure one-sample tests showed no significant deviations from zero. Overall, our results here were not strong enough to support our hypothesis that semantic distance

in the VGT would increase more after a night of sleep than a day of wakefulness in both the random and the creative settings.

Interestingly, overnight change in the VGT showed a possible relationship with time spent in Stage 2 sleep. Importantly, we found no effect of time of day on performance in the VGT, nor an effect of alertness, and alertness and sleep parameters did not differ per session or group. In terms of the validity of semantic distance as a measure of creativity, we show similar results to Heinen and Johnson (2018). Namely, we found that a prompt to be creative led participants to find a middle ground between common and random responses, apparently taking into account both the originality and appropriateness requirements of creativity.

Surprisingly, in the AUT a period of wakefulness, not sleep, led to more and more creative ideas about uses for everyday objects. Additionally, this over-day change was significantly higher than zero, whereas the overnight change was not different from zero. When separated by measure, overnight and over-day change showed a trend difference in the fluency and flexibility measures, while originality was non-significant. In all these measures, over-day change was significantly better than zero, while overnight change did not differ from zero. Notably, though, the AUT results were confounded by a time of day effect where participants scored higher in the evening than in the morning. Higher alertness, as measured by PVT reaction time, was also related to higher performance on originality in the AUT. We did not find any significant relationship with sleep parameters. Importantly, alertness and sleep parameters did not differ per session or group. Finally, we found that performance on the AUT and semantic distance in the VGT were not correlated with each other.

Before discussing the results of this chapter further, it is important to point out that the limited sample size used has likely led to the achievement of low statistical power. This was noted in the methods section, but deserves to be pointed out again. In other words, although the results presented and discussed here provide very interesting avenues for further research, future studies working with these tasks should make it a priority to collect a much larger amount of participants in order to draw reliable and replicable conclusions.

4.5.1 Sleep leads to higher semantic distance in the VGT than wakefulness

The relationship between sleep and semantic memory has been studied quite extensively, for instance using the Deese-Roediger-McDermott (DRM) paradigm (Roediger & McDermott, 1995).

This task presents participants with lists of words that are closely related (e.g. flour, toast, loaf, sandwich, crust, etc.), and tests them on their memory for old words, new words, and so-called lure words. Lures are words that were closely related to the words in the list, but that did not actually appear in it (bread, for the list above). Several studies have found increased false memories, i.e. acceptance of lure words as having been part of the word list, after a period of sleep compared to wakefulness (Diekelmann et al., 2010; Monaghan et al., 2017; Newbury & Monaghan, 2019; Payne et al., 2009), though this does appear to depend on the characteristics of the memory task used (Newbury & Monaghan, 2019). This increase in false memories has been interpreted as spreading activation in semantic memory from presented word representations to related concepts.

A creative task that depends largely on participants' semantic network is the remote associates test (RAT; Mednick, 1962), which requires finding the solution that is related to three words which have been given (e.g. way, board, sleep; solution: walk). Two studies combining sleep and the RAT have found results that are consistent with this sleep-enhanced spreading of activation across a semantic associative network (Cai et al., 2009; Sio et al., 2013). Particularly, Sio and colleagues found that sleep improved performance on difficult but not easy RAT problems (2013). Because difficult problems require activation to spread farther along the semantic network from the stimuli to the solution word, this was interpreted as evidence in favour of the spreading activation account.

In the current study, we found that semantic distance between a noun stimulus and a verb response was increased more after a night of sleep than a day of wakefulness when the stimulus had been seen before (old condition). Separation of the different outcome measures showed that this was particularly the case when participants were asked to be creative, though the difference between overnight and over-day change was not significant after multiple comparisons correction in that case.

These results are consistent with the idea that sleep boosts spreading activation in the semantic network. Under this account, the words seen before sleep were reprocessed during the night, resulting in spreading activation and an increase in the semantic distance between those nouns and the corresponding verbs participants generated the next morning. In line with previous experiments, where a wake interval did not increase performance on the RAT or false memories in the DRM paradigm (Cai et al., 2009; Monaghan et al., 2017; Payne et al., 2009; Sio et al., 2013), a wake interval did not lead to an increase in semantic distance between nouns and verbs in this experiment. Thus, if reprocessing of the task words occurred during wakefulness, this did not appear to have led to spreading activation and a subsequent increase in performance on the VGT.

Another notable observation is the fact that the common measure, where participants had to generate a verb that was highly related to the noun, showed the same tendency as the other measures (albeit non-significantly). This is somewhat unexpected, given the fact that the objective in this instance was to obtain a low semantic distance. Perhaps the spreading activation made semantically more distant words more easily accessible, even in cases where this was not advantageous.

An alternative explanation for these VGT results is that an interval of wake creates worse performance, rather than sleep being specifically beneficial. This is perhaps more in line with the results of our one-sample tests, which indicate that the semantic distance change over sleep is not significantly higher than zero. Admittedly, this same change over wakefulness was also not significantly worse than zero, but did reach trend level.

Such an interpretation of the results could be explained from the point of view of the synaptic homeostasis hypothesis (SHY; Tononi & Cirelli, 2014). In this view, wakefulness is filled with learning which depends on strengthening connections in the brain. The results of these strengthening connections are that synapses become saturated, eventually reducing our ability to learn. Sleep, then, provides a 'reset' of these connections, by renormalizing synaptic strength throughout the brain and thereby restoring cellular homeostasis. This could provide an explanation for our finding that an interval of wake seems to be detrimental to performance on the VGT, more than a period of sleep being beneficial. From this point of view, however, it is also interesting that an over-day interval does not appear to be disadvantageous for performance on the AUT – on the contrary, it seems to be beneficial.

4.5.2 An over-day interval improves performance on the AUT

As far as we are aware, no study has compared overnight and over-day change in performance on the AUT, despite it being one of the most-used creativity tests (Arden, Chavez, Grazioplene, & Jung, 2010). On the other hand, research looking at the effect of a wake incubation period on creative performance has made extensive use of the AUT. In fact, it has been shown that divergent tasks in particular (in contrast to convergent thinking tasks) benefit from a period of wake incubation (Sio & Ormerod, 2009). Even though these incubation periods are usually only a short break, our over-day interval could be considered a period of wake incubation. Thus, our finding that an over-day interval improved performance on the AUT is in line with previous studies in this regard.

Improvements on the AUT after incubation have been shown using breaks filled with idea evaluation (Hao et al., 2016), mind wandering (Baird et al., 2012), and focussed and unfocussed suppression of the task (Gilhooly, Georgiou, Sirota, & Paphiti-Galeano, 2015), among others. Of particular interest here is an experiment that evaluated the use of tasks that were supposed to stimulate remote associative processes during the incubation period (Hao et al., 2014). In this study, participants worked on the AUT, had an incubation period during which one of four different tasks was completed, and afterwards resumed their work on the AUT again. Of the four incubation-period tasks, two had been related to associative processes and brain activity patterns, and the other two were verbal control tasks which were thought not to elicit these associations. The results showed that performance on the AUT was enhanced after incubation with the associative tasks, but not the control tasks.

Given this study, we might infer that incubation-related improvements are due to remote associative processes that occur during the rest period. Indeed, a prominent theory explaining these improvements is the unconscious thought theory (UTT), which posits that unconscious thought may be more associative and divergent, leading to more original ideas (Dijksterhuis & Meurs, 2006; Dijksterhuis & Nordgren, 2006). This is in contrast with conscious thought, which is considered to be more focused and convergent. The results of several experiments are in line with this idea. For example, Dijksterhuis and Meurs showed that participants who were distracted for a few minutes after receiving the instructions to divergent thinking tasks (unconscious thought condition) produced more original elements than those who immediately completed the task or were instructed to consciously think about it (Dijksterhuis & Meurs, 2006). Other studies have used checks for intermittent conscious work or varied the nature of the creative and incubation-period tasks to support the role of unconscious thought in incubation-related improvements (Gilhooly, Georgiou, & Devery, 2013; Gilhooly, Georgiou, Garrison, Reston, & Sirota, 2012). In light of these results, however, it remains surprising that the ultimate period of unconscious thought – sleep – did not improve performance on the AUT. Indeed, if we consider sleep to elicit remote associative processes, like the spreading activation concept mentioned above would suggest, then we would have expected overnight performance improvements on this task.

In other words, being awake during the incubation period appears to be critical for performance improvements on the AUT. This could point to some involvement of conscious thought, whereby participants continue to think about and work on the problem during the incubation period (Browne & Cruse, 1988; Posner, 1973). It certainly seems possible that participants encountered

objects or situations throughout their day which reminded them of the task, either consciously or unconsciously (for instance “Hey, I guess I could also use a tin to roast food in the oven” while making dinner). This would have been less likely to happen when participants went to sleep shortly after the evening session in the lab. For one, the lack of a benefit of sleep does appear to argue against the idea that incubation effects occur by reducing mental fatigue (Posner, 1973), as participants should have been somewhat more rested after just having slept compared to just having spent the whole day awake.

A related issue is the fact that performance on the AUT was affected by time of day, where participants scored higher in the evening than in the morning. Time of day, together with an individual’s circadian rhythm, has been shown to influence a variety of cognitive processes, for instance problem solving (May, 1999; Wieth & Zacks, 2011), memory (May, Hasher, & Foong, 2005; May, Hasher, & Stoltzfus, 1993), and attention (Intons-Peterson, Rocchi, West, McLellan, & Hackney, 1998). Although we did not measure morningness and eveningness in our participants, it has been shown that young adults (our sample) tend to be evening or neutral types, rather than morning types (May et al., 1993). How this may affect their performance on the AUT can only be speculated, given the range of cognitive processes that could affect this task in different ways. For example, explicit memory retrieval is better at participants’ peak time of day (May et al., 2005), and this may in turn allow participants to come up with more or more original alternative uses for an object. On the other hand, it has also been shown that reductions in attention, which were associated with non-optimal times of day, led to higher rates of insight problem solving (Wieth & Zacks, 2011). This indicates that performance on (some) creative tasks might be better when time of testing does not align with participants’ preference. Given the fact that the current study design does not allow us to separate effects of sleep and wakefulness from circadian effects, it might be interesting to look at AUT performance in an experiment using a daytime nap.

4.5.3 Possible explanations for the disparity between the AUT and VGT

While the AUT and the VGT clearly share some aspects, our results show that they are not influenced in the same way by overnight and over-day intervals, and time of day. In a previous study, the creative part of the VGT was shown to be strongly related to AUT performance; a relationship which the authors used in their validation of the creative version of the VGT (Prabhakaran et al., 2014). However, we did not find this correlation in our experiment, and indeed this lack of correlation is in line with the difference in performance change over the different intervals. Without a doubt, the AUT is a more complex task than the VGT, making use not only of

participants' semantic network, but also their episodic memory and their ability to imagine and mentally manipulate and deconstruct items (Gilhooly et al., 2007). In other words, intervals involving wakefulness and sleep may be influencing different aspects of the tasks.

Additionally, the AUT is a task which has been known to depend on working memory capacity (Hao, Yuan, Cheng, Wang, & Runco, 2015) and prefrontal cortex functioning (Vartanian et al., 2014, 2013). Certainly, the AUT demands more sustained attention, given that the task continued for three minutes per object. In contrast, the VGT is a much quicker task, which may depend more on automatic processes. In the VGT, participants could move the task along by themselves as soon as they had filled in a verb response, and it is likely that many did so to finish the experiment quickly. Thus, perhaps the VGT is more susceptible to the associative processes that are thought to happen during sleep, whereas the AUT allows for more involvement of slower processes (e.g. mental imagery) and executive control. Indeed, the VGT and AUT may still overlap in some way, but whatever is responding to sleep in the VGT does not appear to determine performance on the AUT. Perhaps the processes affected by sleep are a sufficiently small part of completing the AUT, and as such they are drowned out by other factors that determine the performance there. This may also tie in with the fact that we found an effect of time of day in the AUT, but not in the VGT. This lack of an effect of time of day in this task is in line with the idea that the VGT depends more upon automatic processes, given that there is some evidence that automatic retrieval processes do not differ throughout the day (Yang, Hasher, & Wilson, 2007).

4.5.4 Relationships with alertness and sleep parameters

Differences between the AUT and VGT may also be rooted in their relationships to sleep and alertness. In the VGT, we did not find a correlation between alertness and task outcomes. In contrast, in the AUT, we found a negative correlation between originality and reaction time on the PVT. Thus, in this task, when people were more alert (they showed a lower reaction time on the PVT), they came up with more original responses. Given that alertness (as measured by both the questionnaires and the PVT) did not differ between the morning and the evening, this points again to the possibility of an inter-individual time-of-day component. Some participants were more alert in the morning, and others in the evening. Thus, the optimal time to be creative may differ per person. Nevertheless, the relationship with alertness does not appear to be very strong, given that only the originality measure showed this association.

The relationship with alertness, however small, is not surprising. Several studies have found that sleep deprivation (presumably leading to lower alertness) impairs creative performance (Horne, 1988; Vartanian et al., 2014; Wimmer, Hoffmann, Bonato, & Moffitt, 1992). In this regard, it may be interesting to note that convergent thinking tasks appear to be more resilient to short-term sleep loss than divergent thinking tasks (Horne, 1988). The VGT, particularly the creative measure, contains a significant convergent aspect. Namely, although participants are allowed to think of many different verbs, they must choose the most appropriate one to fit the instructions. As mentioned above, the VGT shows no relationship with alertness. On the other hand, performance change on the AUT, a highly divergent task, was correlated with PVT RT in the originality measure. In other words, the difference in the relationship with alertness, potentially relating to circadian confounds, may be another reason for the disparate findings between the VGT and the AUT.

In the correlational analysis with sleep, Stage 2 came forward as a candidate for involvement. In the VGT, the random measure showed a positive correlation with time spent in S2, whereas we did not find significant correlations with sleep in the AUT. The connection with Stage 2 sleep is intriguing, given that this is the stage during which sleep spindles are most prevalent (De Gennaro & Ferrara, 2003; Fernandez & Lüthi, 2020). Sleep spindles have long been considered to play an important role in memory consolidation (e.g. Astori, Wimmer, & Lüthi, 2013; Cairney, Guttesen, El Marj, & Staresina, 2018; Fogel & Smith, 2006; Jegou et al., 2019; Schabus et al., 2004; Ulrich, 2016), which may in turn aid creative processes. Surprisingly, we did not find a significant relationship with REM sleep, which may have been expected because it has been implicated in past creativity studies (Cai et al., 2009; Lacaux et al., 2019; Walker, Liston, et al., 2002).

While this experiment has highlighted interesting effects of an overnight interval on creative performance, particularly in the VGT, one thing that remains difficult to pinpoint is the precise involvement of sleep and different sleep stages. First, because participants slept at home with a headband and not in the lab with polysomnography, which would give a more accurate picture of participants' sleep. Second, this study looked at overall effects of sleep and wakefulness, and did not manipulate sleep in any way. A great next step would be to conduct a similar study using targeted memory reactivation (TMR), where sounds or smells are paired with task items and re-presented during sleep. An obvious sleep stage for TMR would be REM sleep, given its purported links with creativity, but results from this experiment suggest that such a TMR study should perhaps also include reactivation in Stage 2 sleep. Although we did not find any beneficial effects of an overnight interval on AUT performance, manipulations during sleep (i.e. TMR) might expose these.

An interesting idea to consider is the possibility that this task may still benefit from sleep, but any impact of this in our experiment was occluded by the time of day effect. As mentioned earlier, a nap study may be better suited to distinguish the effects of wakefulness and sleep on the AUT.

4.5.5 Conclusion

The aim of this study was to examine the effects of an overnight and an over-day interval on two creative tasks: the VGT and the AUT. We demonstrate that change in semantic distance score on the VGT was higher after an interval containing sleep compared to an interval containing wakefulness. This may have been caused by a sleep-related increase of spreading activation in participants' semantic networks. However, given the fact that change in semantic distance was not significantly higher after an interval of sleep, the difference between our overnight and over-day intervals could also have come from synaptic saturation which may occur during wakefulness. This would have reduced performance on the VGT at the end of the day, with sleep allowing synapses – and subsequently next-day performance – to recover. In the AUT, on the other hand, we found that an over-day interval was more beneficial for performance than an overnight interval, although this was confounded by a time of day effect. The over-day benefit fits with previous findings regarding the effect of wake incubation periods on AUT performance and divergent thinking tasks more generally. The disparity between the VGT and AUT results may be explained by the more complicated nature of the AUT, performance on which is influenced by several different strategies and processes, in contrast with the VGT which depends highly on semantic processing. Although the precise involvement of sleep and sleep stages in creativity remains ambiguous, this experiment showed that sleep may indeed benefit certain creative tasks.

CHAPTER 5

General discussion

5.1 Overview

The aim of this thesis was to add to the current understanding of the role of sleep in the reprocessing and restructuring of memory. Nowadays, there is no question that sleep can benefit memory consolidation. However, memories are not carbon copies of our experiences; they are abstracted, integrated, and restructured into complex networks. These networks allow us to understand our environment, make connections between experiences, and adaptively deal with situations we may not have encountered before. There is evidence that sleep benefits restructuring processes as well.

However, this does not happen for all memories, nor under all circumstances. Moreover, sleep consists of different stages, all with different properties – whether and how these different stages are involved remains unclear. Beyond that, there is also still much to learn about the mechanisms of memory restructuring during sleep. This process is thought to be driven by the reprocessing (or reactivation) of memories, but precisely which of our myriad of memories are reprocessed and restructured? And how do sleep and wakefulness compare when it comes to memory restructuring?

These questions inspired the research conducted throughout this thesis. To address them, I have used polysomnography, targeted memory reactivation, comparisons of wake and sleep, and several different tasks. In chapter 2, I focused on memory reactivation during the serial reaction time task. This is a task we know quite a lot about, but the results of my experiment show that there is still much we can learn from it. In chapter 3, I looked to the associative inference task, which had not been explored with TMR before. Lastly, in chapter 4, I turned my attention to the exciting field of creativity, and compared effects of wake and sleep on two creative tasks.

I will begin this final chapter by summarising the findings arising from each of the experiments of this thesis, and discussing their limitations. Then, I will integrate the results of these experiments with previous literature, first looking at memory reprocessing. What can my experiments and previous research tell us about the roles of SWS and REM in memory reprocessing? And what do we know about the selectivity of this reprocessing? Subsequently, I will further link my experiments by turning our attention to memory restructuring. In light of my findings, can we say that sleep indeed promotes memory restructuring? Is this memory restructuring also selective, and how does sleep compare to wakefulness regarding its influences on the reorganisation of memory? Finally, I

will link memory reprocessing and restructuring explicitly, and discuss important remaining questions that may direct future research.

5.2 Experiments in this thesis

5.2.1 Summary of findings

Chapter 2 had three complementary aims. First, we wanted to investigate whether TMR of a serial reaction time task in SWS and REM would lead to implicit and explicit memory restructuring. Second, we were interested in whether this would be equally the case for both hands, which might tell us more about the selectivity of memory restructuring. Finally, we also wanted to examine whether we could detect memory reactivation in both SWS and REM, to find a concrete link between memory reprocessing and restructuring. We found that SRTT performance benefitted from TMR in SWS, but not REM. In line with these findings, our machine learning classifier was able to reliably detect memory reactivation during SWS, but not REM sleep. Furthermore, SWS TMR significantly improved sequence performance in the non-dominant, but not the dominant hand. Together, these results confirm and extend the importance of SWS in memory reactivation and restructuring. Importantly, we did find some evidence of memory processing during REM sleep, namely an ERP difference between cues related to the left and right hand, and marginally significant memory reactivation detection with our classifier after removal of one outlier. However, given the lack of behavioural improvements in the REM group, the functional relevance of this processing during REM is unclear. Finally, the fact that we only found a significant behavioural improvement in the left hand may indicate conditions under which TMR thrives: perhaps TMR works best for weaker memories, or for memories which are processed more bilaterally in the brain.

In Chapter 3, we investigated memory reprocessing and restructuring with an associative memory task. We were interested in the behavioural effects of TMR during SWS and REM, both on associations that were learned before sleep, and on novel associations that had to be inferred. Previous research had indicated that sleep was beneficial for both of these associations, but there was uncertainty about the influence of different sleep stages (Alger & Payne, 2016; Lau et al., 2010). Although we hypothesised that TMR during SWS would improve direct associations, we did not find this to be the case, potentially because participants' performance was already too high before sleep. Our findings in the REM group, on the other hand, were in line with our expectations. Participants showed higher accuracy on items in the inference task that had been cued during sleep compared to uncued items, which indicates a role for REM sleep in memory restructuring.

Furthermore, ERP results demonstrated processing of TMR cues during REM sleep. However, two replications of the REM group, with new participants, did not reveal the same behavioural or electrophysiological effects. These group differences may have been due to individual differences in TMR susceptibility, potentially mediated by task structure, or the behavioural effect in the first REM group may simply have been a coincidence. Either way, the three successive experiments outline the importance of careful replications, and raise questions about the robustness and individual differences of TMR that have thus far not been thoroughly addressed in the literature.

In the final experimental chapter, Chapter 4, we wanted to establish the effects of wakefulness and sleep on two creative tasks. One of these tasks, the verb generation task, relies heavily on participants' semantic memory networks, and may thus benefit from sleep-related restructuring of those networks. The other task, the alternative uses task, was chosen because it is one of the most-used tasks in the creativity literature. Moreover, this task draws on various memory systems, most notably episodic and semantic memory, and was thus also a good candidate for benefits associated with memory restructuring during sleep. In the VGT, change in semantic distance between given nouns and generated verbs was indeed higher after an overnight interval than an over-day interval. Thus, sleep may increase the spreading in semantic memory networks, or restructure them in such a way that allows participants to come up with semantically more distant words. An alternative explanation here focused on the idea that sleep may have functioned as a 'reset' of the synaptic potentiation that happened during wake. Surprisingly, we found that performance on the AUT was improved after an interval containing wake rather than sleep, although this was confounded by a time of day effect where participants performed better in the evening than in the morning. These results indicate that sleep does not benefit creativity indiscriminately, and that complex tasks such as the AUT may be more strongly influenced by other processes.

5.2.2 Limitations of the experiments

Power

In all experimental chapters of this thesis, as in the field of sleep research in general, the issue of statistical power is relevant. In my view, a lack of power is one of the main limitations of not only my own work presented here, but of most sleep studies. In short, statistical power has been defined as the probability that your study will detect an effect of interest, given that this effect actually exists in the population you are studying (Cohen, 1962). The level of power generally considered acceptable is 80%, meaning that you will have an 80% chance of obtaining a statistically significant

result for a true effect of interest (Brysbaert, 2019). However, as noted in Chapter 3, an underpowered study not only decreases the chance that you will find a true effect (a false negative), it also increases the chance that you will find an effect that does not actually exist (a false positive) (Button et al., 2013; Fraley & Vazire, 2014). Both of these issues are relevant for the experiments presented in this thesis.

In Chapter 2, a post-hoc power calculation indicated that we may have been sufficiently powered for the large effect that TMR during SWS generally has on the SRTT. Indeed, I do not doubt the presence of a true effect of SWS TMR on performance in the SRTT, given that it has been replicated a number of times now (Cousins et al., 2014, 2016). Nevertheless, it has been shown that studies with small samples tend to overestimate effect sizes (Brysbaert, 2019; Maxwell, 2004), which makes it likely that our statistical power was still on the low side even in this study. Certainly, this experiment was underpowered for medium and small effects. For instance, it is quite possible that we would have found a significant effect of SWS TMR in the dominant hand (in addition to the significant effect we already found in the non-dominant hand), had the number of participants and consequently the statistical power been higher.

The issue of power was mentioned several times in Chapter 3, in relation to the fact that we decided to collect two additional participant groups and in reference to the combined analyses that were conducted. Individually, the experiments in Chapter 3 were likely underpowered for the medium-sized effect we thought we had found. Thus, individually, the significant effect of the first experiment in this chapter and the null effects of the second and third experiments could all have been wrong. However, from the higher-powered combined analyses, we concluded that TMR during REM sleep does not reliably benefit performance on our associative memory task.

Chapter 4 contains two different tasks with corresponding effect sizes. From the post-hoc power analyses, it is clear that this study especially suffers from a lack of power. This is likely the cause of at least the null results in the separate analyses for each measure in the VGT. More concerning, though possible, is the idea that some of the significant findings are actually false positives. Thus, although the results in this chapter provide very interesting avenues for further research, future studies working with these tasks should make it a priority to collect a much larger amount of participants in order to draw reliable and replicable conclusions.

Adding participants is only one way of increasing the power in your study. For labour-intensive experiments like sleep studies, where the collection of one participant generally takes at least an

entire night, adding more participants can be very difficult. However, especially in these occasions, it remains important to conduct high-powered studies. After all, it would be a shame to spend all that time and effort on an experiment that does not offer much insight (Brysbaert, 2019). Other aspects of a study that can increase statistical power mainly concern the design and theoretical basis of the study. For instance, using repeated measures and within-participant (rather than between-participant) comparisons make attaining a reasonable amount of power easier (Brysbaert, 2019). Additionally, expanding the number of observations per participant per condition is a good way to reduce noise and thereby increase power in a study (Brysbaert & Stevens, 2018). It is reassuring that the experiments presented in this thesis are already making use of these strategies. Besides the implementation of these measures, however, it seems clear that sleep studies would benefit from increasing the 'standard' number of participants that are tested. As others have said before, this would require a change in the way that this research is evaluated (Brysbaert, 2019). For example, for a PhD thesis in the field of sleep science, it is generally expected that the student will complete 3-5 experiments (depending on the complexity of each experiment). Perhaps we should require 1-2 well-powered experiments instead.

Other limitations

Beyond issues relating to power, the work conducted for this thesis had some other limitations. In Chapter 2, the main limitation was the inclusion of the motor imagery task. In this task, participants were instructed to follow the sequence of images on the screen like normal. But, rather than pressing a corresponding button in response to each image, they were asked to *imagine* pressing these buttons. This task was added for the benefit of the classifier, because we hypothesised that the EEG data collected during this task would be comparatively free from movement artifacts. A drawback of this task, however, was that it added about an hour of extra time to the experiment (half an hour in the evening and half an hour in the morning). Participants thus had ample time to study both the cued and uncued sequence in a task that required relatively little else of them. It is therefore probable that participants knew both sequences very well, which may have been the cause of the null findings in the explicit memory test (see supplementary analyses, page 75). Although this was an unintended consequence of the inclusion of the imagery task, the intended benefits to the classifier appear to have been achieved, judging by the successful classification of memory reactivation during SWS. In other words, the inclusion of this task limited our ability to study the true effect of TMR during SWS on the explicit memory of a sequence. Nevertheless, it did allow us to achieve one of the main goals of the study.

The limitations of the experiments in Chapter 3 have been discussed before, but they deserve reiterating. Most notably, there is the issue of our manipulation of the gender of the probe face in the remote (face-face) associations test. In the first experiment of this chapter, this gender was kept constant, but in the second experiment it was switched around between sessions. This appeared to have affected participants' performance, and may have been the cause of the null findings in the second experiment. This issue represents one of the main things I would have done differently with hindsight, as it could have prevented the need for a second replication of the REM group (i.e., Experiment 3 in Chapter 3). Another limitation in this chapter, further discussed both in section 5.3.1 below as well as in the discussion of Chapter 3, is possible ceiling effects leading to null results in the SWS group. Because performance on the learned associations was so high from the start, TMR during SWS likely served little additional benefit. Thus, we are limited in the conclusions we can draw about the effect of SWS TMR on this task.

Lastly, the limitations of Chapter 4 are mainly concerned with the design of the study. As mentioned in the discussion of this chapter, one difficulty is that we cannot pinpoint the precise involvement of sleep and different sleep stages. This is due to the fact that participants slept at home with a headband (rather than in the lab with polysomnography), and because we did not manipulate sleep in any way. As a result, we can only make very general statements about a possible effect of sleep and we have to rely on correlations to indicate whether particular sleep stages might be related to the behaviour we observed. A related issue with the design is the fact that it does not allow us to fully separate circadian effects on the tasks. Although we were able to measure the effect of time of day by analysing the new items completed at each time point, when this analysis revealed a time of day effect in the AUT we could not be sure to what extent this circadian confound altered the effects of sleep and wakefulness. Thus, we recommended a follow-up experiment using a daytime nap. In such an experiment, we would also recommend including measures of participants' circadian rhythms, like the Morningness Eveningness Questionnaire (Horne & Östberg, 1976). This would allow us to analyse how participants' peak time of day relates to their performance on the tasks. In turn, this could for example shed more light on potential reasons for the benefit of wakefulness we observed in the AUT.

5.3 Memory reprocessing during sleep

5.3.1 Reprocessing during SWS

In the introduction, I reviewed a wealth of research which indicates that memory reactivation is the basis of sleep's role in memory. Most of this research, and thus most of the evidence showing memory reprocessing during sleep, has looked at NREM or SWS. The studies are too numerous to review again here, but I will provide a quick overview with evidence from three different angles. Ever since TMR has gained renewed popularity, there has been abundant indirect evidence for memory reprocessing during SWS. Very specific behavioural effects of TMR have been found, down to the individual memory (e.g. Rudoy, Voss, Westerberg, & Paller, 2009). Additionally, in rodents, there has been direct evidence of memory reactivation during sleep since the 90s (Skaggs & McNaughton, 1996; Wilson & McNaughton, 1994). Importantly, two studies in humans have now shown, using similarity analyses and classifiers, that learning-related activity is reinstated following the presentation of an auditory cue (TMR) during NREM sleep (Belal et al., 2018; Schreiner et al., 2018).

Our results from Chapter 2 are clearly in line with the evidence presented above and in the introduction. Not only did we find behavioural effects of TMR during SWS, but we were also able to detect memory reactivation during sleep in this stage. Specifically, we were able to distinguish cues related to left and right hand button presses, and when we compared this classification during the experimental night to that in the adaptation night (when the sounds had no memory component yet), it was significantly higher. Although we did not find a correlation between classification strength and behavioural improvements, we did find that the extent to which SWS TMR-elicited ERPs were greater in the experimental compared to the adaptation night was negatively associated with TMR-related performance improvements in the left hand after sleep. This provides some link between electrophysiology and behavioural effects.

By comparing classification during the experimental and adaptation nights, we were able to show that this method detects the memory related to a sound, and not merely a differing electrophysiological response to different sounds. This had not been directly demonstrated in previous studies, though various controls were employed which indicated the same (Belal et al., 2018; Schreiner et al., 2018). In other words, Chapter 2 provides additional support for the fact that memory reactivation takes place during SWS, and extends previous findings with novel methodology and rigorous controls.

The picture becomes more complicated when we take into account our findings from Chapter 3. Behaviourally, we did not find any effect of TMR in SWS on either learned or inferred associations. We also did not use a classifier in this experiment, mainly because the design of the task (particularly the 60 different sounds used) would have made its implementation very difficult. On the other hand, we did find some indications of memory reprocessing in the time-frequency results of the SWS group. Namely, we found a fast spindle band response after experimental but not control sounds, although the difference between these was not significant and did not correlate with behaviour.

At face value, it seems difficult to reconcile the lack of behavioural results in the SWS group of Chapter 3 with previous studies indicating that sleep does benefit both learned and inferred associations (Alger & Payne, 2016; Lau et al., 2010), especially considering that one of the studies specifically found a relationship with SWS (Lau et al., 2010). However, we must keep in mind that those studies looked at equivalent periods of sleep and wakefulness, whereas we looked at equivalent periods of sleep manipulated with TMR. Indeed, a closer look at the learned association results in our experiment shows that evening and morning accuracy was comparable, indicating that participants maintained performance overnight – but there was no additional benefit of TMR. In the previous studies, performance did decrease from baseline to after the intervening period, where participants in the nap group forgot less than those that stayed awake (Alger & Payne, 2016; Lau et al., 2010). As mentioned in Chapter 3, we suspected that participants were overtrained in our experiment, which might have obscured any effects that TMR and associated memory reprocessing could have had. The fact that previous studies showed forgetting over an interval spanning 3-4 hours, but our experiment did not show forgetting overnight, indicates that participants remembered the learned associations very well.

It may also be interesting to note our findings regarding the time course of memory reactivation during SWS. Previous research has indicated that the period of maximal decodability of reactivations was around 2 seconds after cue onset, during a fast spindle band increase in response to the cues (Cairney et al., 2018). However, in Chapter 2, we found that the period of highest classification in the SWS group was around 1 second after cue onset. This difference may be related to the fact that subsequent cues were presented approximately 1.5 seconds after one another in our experiment. It is possible that the rapid succession of cues in our experiment stopped evoked memory reactivations prematurely, although our behavioural TMR benefits did not appear to be negatively affected. Nevertheless, in light of the study by Cairney and colleagues (2018), which

came out after the data for Chapter 2 had been collected, it would be very interesting to investigate how the behavioural effects and time course of classification might change if we presented cues further apart from each other. Such an experiment is now underway in our lab.

Similarly, previous research has found that presenting a second cue soon after the first can eliminate the benefits of TMR (Schreiner et al., 2015). While our inter-trial interval in Chapter 2 did not appear to be too short, judging by the behavioural TMR effects we found, one could wonder whether the lack of TMR results in Chapter 3 may have been caused by a similar mechanism. In Chapter 3, sound cues were all 2 seconds long (compared to 200 ms in Chapter 2). This sound length was chosen because pilot tests showed that participants found it difficult to determine the sound identity when the cues were shortened. However, it is possible that ongoing auditory input during these 2 seconds could have blocked emerging reactivation and eliminated potential TMR benefits. However, the time frequency findings, which show a trend difference between experimental and control sounds about 1.2 seconds after cue onset, do not seem to be in line with this idea. Nevertheless, these unresolved questions indicate that the time course of reactivation remains an important area of focus for future research.

Overall, our results from Chapter 2 provide strong evidence that memory reprocessing occurs during SWS, that we can detect it using learned sounds and classifiers, and that its occurrence during sleep provides behavioural benefits. Combined with the results from Chapter 3, we can say that memory reprocessing during SWS may still be caused to occur with TMR even when a strong memory was already established before sleep, as indicated by the time-frequency findings. However, in this instance, TMR may not lead to the benefits so often associated with it. Future research should focus on the conditions under which memory reprocessing occurs, its time course, and on the conditions that allow for behavioural benefits.

5.3.2 Reprocessing during REM

The evidence for memory reprocessing during REM sleep is altogether more scattered. There are several early studies that have seemingly successfully triggered memory reactivation during REM (Guerrien et al., 1989; Smith & Weeden, 1990), and two more recent ones (Rihm & Rasch, 2015; Sterpenich et al., 2014), but there are also several studies that have found no behavioural effects of TMR during REM sleep (Cordi et al., 2014; Laventure et al., 2016; Lehmann et al., 2016; Rasch et al., 2007). Within the rodent literature, there is some direct and indirect evidence for memory reactivation in REM sleep (Howe et al., 2019; Louie & Wilson, 2001; Pavlides & Winson, 1989; Poe

et al., 2000), though not nearly the overwhelming amount that has been found in NREM sleep. In humans, early work using positron-emission tomography showed that blood flow during REM sleep was affected by tasks completed before sleep (Maquet et al., 2000; Peigneux et al., 2003). Finally, one study using multivariate pattern analysis was able to detect learning-related patterns of EEG activity during REM sleep, although this did not correlate with subsequent memory performance (Schönauer et al., 2017). In short, there is some evidence that memories are reprocessed in REM sleep, though much less is known about this processing than that during NREM sleep.

Within this context, it is important to point out that there are certain complications which contribute to the relative shortage of (successful) REM reactivation studies. One of the issues surrounding the detection of memory reprocessing during REM is the fact that it is simply a lot more difficult to clean than NREM sleep. The rapid eye movements which are characteristic of this stage are reflected in the frontal and sometimes even the central EEG channels. While several methods have been devised to remove these, like regression analysis and independent component analysis, these methods are far from perfect (Schlögl et al., 2007; Vigário, 1997). It is likely that some brain activity of interest will be removed, and that some residual eye activity remains even after removal. This would make it more difficult to match up learning-related brain activity during wake and REM sleep. As an added problem, TMR during REM sleep is complicated by a decreased arousal threshold in this stage compared to NREM sleep – especially SWS (Busby, Mercier, & Pivik, 1994; Neckelmann & Ursin, 1993). This threshold is further decreased as the night progresses and homeostatic sleep pressure is reduced, which makes it difficult to stimulate during late REM sleep. In other words, the volume of auditory cues needs to be lower during REM sleep to reduce arousals, but this also reduces the likelihood that participants will hear the TMR cues. However, when they do clearly hear the cues, they are more likely to wake up from them. Several REM TMR studies have therefore used olfactory rather than auditory stimuli, but none of these have been successful (Cordi et al., 2014; Laventure et al., 2016; Rasch et al., 2007). Because the two recent REM TMR studies which resulted in behavioural effects both used auditory cues (Rihm & Rasch, 2015; Sterpenich et al., 2014), and because we know that auditory information is processed during REM sleep (Bastuji & García-Larrea, 1999; Niiyama et al., 1994; Sallinen et al., 1996; Takahara et al., 2006), we chose to employ auditory TMR despite the complications.

The results of Chapter 2 are largely in agreement with the literature. While we did not find a behavioural effect of REM TMR in this chapter, we did find a difference in the ERPs to left and right-handed cues. In addition, we showed that classification of memory reactivation during REM is in

principle possible, although currently not to a level that we trust as meaningful. Together, these are indications that memories are indeed also reprocessed during REM sleep. However, the lack of TMR-related behavioural benefits raise important questions about the relationship between memory reactivation and task performance, and the function of REM reactivation more generally. What is the function of memory reactivation, if not consolidation and restructuring leading to more adaptive behaviour?

An interesting idea to consider is that perhaps the function of REM reactivation is not currently captured by our behavioural measures. This idea would fit somewhat with the sequential hypothesis, which suggests that benefits of sleep are highest when NREM and REM follow each other cyclically (Giuditta et al., 1995). In this interpretation, selective behavioural benefits on this task would depend on memory reactivation during NREM sleep, and (targeted) memory reactivation during REM sleep would be ineffective if it was not preceded by NREM reactivation. Put differently, TMR during SWS would lead to selective performance increases in cued items, potentially further benefitting from subsequent REM sleep. This also fits with results from a previous study using the SRTT, where performance increased with SWS TMR, but task-related brain activity in specific areas was associated with time spent in REM sleep (Cousins et al., 2016). TMR during REM sleep, on the other hand, would only lead to general benefits, because preceding NREM sleep (without TMR) would have already strengthened both cued and uncued items.

Adding in the results from Chapter 3, these may raise more questions than they answer. The first experiment in this chapter provides important new information regarding memory reprocessing during REM, specifically the fact that TMR led to a behavioural improvement and the fact that we found differences between memory-related and control cues in the ERP. However, the results from the replications that followed call these findings into question. It remains difficult to interpret the combined results from these three experiments. The simplest explanation is of course that the initial effects we found were coincidental, and there is no real effect of TMR during REM sleep in this task. However, this seems difficult to defend, given the fact that we found two indications of TMR effects: a significant difference between cued and uncued items in the inference task in the two-week follow-up, and an indication of memory processing in the ERPs. If both of these results were flukes, as the fact that they were both absent from the replications would suggest, their combined occurrence in one experiment is quite coincidental.

A more interesting explanation, which would account for the remarkable differences between the experiments in Chapter 3, has to do with individual differences in susceptibility to TMR. This may

for some reason be particularly the case for memory reactivation during REM sleep, which would explain the null findings that often occur in REM TMR studies. Certainly, the idea of individual differences in sleep and memory reactivation is not controversial. For instance, we know that there are large differences in people's sleep duration and oscillatory activity during sleep (Buckelmüller et al., 2006; Purcell et al., 2017; van Dongen et al., 2005; Werth et al., 1997). One study has further found that participants' working memory capacity was positively associated with sleep-dependent memory consolidation (Fenn & Hambrick, 2012). Yet another study showed that explicit knowledge in the SRTT was related to the amount of transitions between NREM and REM sleep in a trait-dependent way (Kirov et al., 2015). Notably, these types of individual differences are also what allow us to find correlations between cueing benefit and electrophysiological or sleep measures, like in Chapter 2. Admittedly, in Chapter 3 we did not find any correlations between behavioural changes and brain or sleep measures. Nevertheless, there may have been some aspect of the electrophysiology or behaviour that we were not able to measure. However, even taking individual differences into account, it does appear that the effects of REM TMR are not very reliable in this task, judging by the fact that the majority of the experiments in this chapter did not find any behavioural or electrophysiological changes.

In summary, both Chapter 2 and 3 indicate that some memory reprocessing can occur during REM sleep. What the function of this reprocessing is, and whether it is indeed dependent on individual differences, remains to be investigated. It is clear that there are many uncertainties, inconsistent findings, and null-results in studies looking at REM, which makes formulating a cohesive theory very difficult. Adding to this are the difficulties surrounding REM TMR and the detection of reactivation in the EEG during this sleep stage. I believe that the functions of REM sleep, and their relationship with memory reactivation, are some of the key remaining questions in the sleep and memory field.

5.3.3 The selectivity of reprocessing

Another important remaining question is that of the selectivity of reprocessing. Which memories are reactivated, and how is this determined? It has been proposed that initial selection of relevant (and irrelevant) memories already occurs during or shortly after encoding (Stickgold & Walker, 2013). This 'tagging' of memories for reprocessing may be particularly geared towards the selective remembering of information that will be relevant in the future. For example, Wilhelm and colleagues showed that benefits of sleep-dependent memory consolidation depend on whether participants were aware that they would be re-tested after sleep (Wilhelm et al., 2011). In another study, participants learned words that were followed by a cue which explicitly told them to either

remember or forget the word (Saletin, Goldstein, & Walker, 2011). This study showed that a period of sleep after learning, compared to wakefulness, resulted in enhanced recall of words to be remembered, while words to be forgotten were selectively excluded from this sleep benefit. Research has also shown preferential consolidation during sleep of information that is associated with a reward (e.g. Abe et al., 2011; Fischer & Born, 2009; Studte, Bridger, & Mecklinger, 2017), and emotional information (e.g. Hu, Stylos-Allan, & Walker, 2006; Payne, Chambers, & Kensinger, 2012; Sopp, Michael, Weeß, & Mecklinger, 2017). Furthermore, there is evidence that consolidation and reactivation favour memories that were weakly encoded (Cairney et al., 2016; Drosopoulos et al., 2007; Schapiro et al., 2018; Tambini et al., 2017).

Our results from Chapter 2, where we found a significant effect in the left but not the right hand, may be related to this preferential treatment of weak memories. Nevertheless, in the discussion of that chapter we also mentioned competing explanations, such as the more bilateral demands that the left hand places upon the brain. Future experiments using left-handed participants may be able to illuminate this issue. In a way, our results from Chapter 3 may also correspond to the idea that weaker memories benefit more from reprocessing during sleep. We suspected that participants' memories for learned associations were too strong to be affected by TMR. Correspondingly, we found no effects of TMR on these associations. However, as a recent preprint has noted, these and similar experiments suffer from ceiling effects which may obscure sleep effects on stronger memories. In a new memory paradigm, they showed that stronger memories also benefit from sleep under certain conditions (Petzka, Charest, Balanos, & Staresina, 2020, PsyArXiv). This indicates that all memories undergo consolidation (through reactivation) during sleep, but for stronger memories this may only be reflected in next-day behaviour when testing demands are increased. These results are in line with our findings from Chapter 3, given that there was likely a ceiling effect there, but how this relates to our procedural task from Chapter 2 remains to be seen.

In all experimental chapters of this thesis participants were aware that they would be retested after sleep or in subsequent sessions, which should ensure that what they were learning had future relevance (Wilhelm et al., 2011). However, in Chapter 3 we purposely did not inform participants that they would be tested on remote associations, and these had also not been explicitly learned. Previous research has indicated that awareness of learning a skill, as distinct from awareness of retesting, additionally modulates the benefits of sleep (Robertson, Pascual-Leone, & Press, 2004). Thus, the fact that this remote inference skill was not explicitly learned may have contributed to the lack of TMR benefits on remote associations. On the other hand, although awareness may

indeed have been a factor, sleep benefits on learning have also been shown in implicit tasks (Cousins et al., 2014; Durrant et al., 2011). One must also remember that the objective of the study was to investigate novel associations and memory restructuring. If we had told participants to anticipate being tested on these remote associations, they would not have required memory restructuring during sleep to complete this.

As a final thought, one may wonder whether experiments using TMR are an appropriate way to determine the selectivity of memory reprocessing. It is possible that TMR actually disrupts endogenous selection mechanisms, given that our cues tell the brain what it needs to remember. Would those same memories have been reactivated, had we not triggered them? However, what is clear from both the literature and from the experiments in this thesis, is that many factors work together to determine which memories are selected for reprocessing and which memories benefit from this. Some of these factors are discussed above, but there are likely many others and no one factor can explain the existing research. This remains an important area in need of more research, because it is at the heart of the link between memory, memory reactivation, and sleep.

5.4 Memory restructuring during sleep

5.4.1 Does sleep really promote memory restructuring?

One of the more intriguing functions that have been ascribed to sleep, which has been an overarching theme in this thesis, is the restructuring of memory. In the introduction I reviewed a number of studies that have indicated that sleep is indeed beneficial for qualitative memory changes. For instance, an abundance of papers has looked at regularity abstraction and generalisation after sleep. Notable examples include generalisation of category learning (Friedrich et al., 2015; Graveline & Wamsley, 2017; Sandoval et al., 2017), and the abstraction of sequence knowledge which leads to increased performance in the serial reaction time task (SRTT) (Cousins et al., 2014, 2016; Maquet et al., 2000). I also examined several papers looking at associative inference, particularly two studies that showed an effect of sleep on making relational associations (Alger & Payne, 2016; Lau et al., 2010). Finally, studies on sleep and creativity were discussed, showing beneficial effects of sleep on (among others) analogical transfer (Monaghan et al., 2015), the remote associates task (Cai et al., 2009; Sio et al., 2013), and coming up with creative solutions (Ritter et al., 2012). Together, all these studies appear to argue quite convincingly that sleep promotes memory restructuring.

However, there are also a number of studies that have shown no effect of sleep on restructuring tasks. For example, in the SRTT, when a probabilistic rather than a fixed sequence is used, no sleep-related benefits are found (Nemeth et al., 2010; Song, Howard, & Howard, 2007). Several category generalisation studies have also found that this ability was not affected by sleep, contrary to the experiments mentioned above (Maddox et al., 2011; Werchan & Gómez, 2014). The lack of sleep effects in some creativity studies has already been mentioned several times (Brodt et al., 2018; Hołda et al., 2020; Schönauer et al., 2018). Thus, the effect of sleep on restructuring does not seem to be universal, and the specific task that researchers choose does appear to affect the outcomes (Lerner & Gluck, 2019).

Indeed, from the experiments presented in this thesis, we cannot draw the conclusion that sleep unequivocally improves memory restructuring. In Chapter 2, we found a TMR effect on implicit (reaction time) learning, but not explicit sequence memory. This implicit effect may be explained by a selective strengthening of the cued items in memory, not necessarily their restructuring. Explicit sequence knowledge would have been a better indicator of memory restructuring, and previous research has indicated that such knowledge can be gained overnight (e.g. Cousins et al., 2014; Diekelmann, Born, & Rasch, 2016; Wagner, Gais, Haider, Verleger, & Born, 2004). Unfortunately, we suspect that participants in our study gained this explicit knowledge during training, particularly during the imagery tasks which were added for the benefit of the classifier and which were not included in previous studies. This does illustrate that such restructuring is likely not exclusive to sleep.

One of the experiments in Chapter 3 shows a positive effect of TMR during REM sleep on remote associations. These results provided strong evidence that sleep promotes novel associations between items that were not learned together, i.e. restructuring of the memory representations. However, the beneficial effects of TMR during REM sleep in this chapter did not appear to be strong enough to come out consistently, as discussed before. Chapter 3 also highlights the importance of task elements on memory consolidation. We found that switching the gender of the probe face in the remote (face-face) associations task likely affected participants' performance. Although it is still unclear whether this had any interaction with the effect of TMR or indeed led to any changes in memory restructuring, it is clear that minor changes to the task did make a big difference to participants' accuracy over time. This is an important consideration that future studies must take into account.

Lastly, Chapter 4 shows performance increases in the VGT after an interval containing sleep. This has been interpreted in the light of spreading activation, leading to novel connections between memory traces that are similar or associated, making more distantly related concepts more easily accessible (Landmann et al., 2014). Thus, under this interpretation the VGT results provide some evidence that sleep is involved in a reorganisation of the semantic memory network. The alternative explanation offered for this finding was related to the fact that the change in semantic distance after an interval of sleep was not higher than zero, only significantly higher than the change after an interval of wake. Thus, perhaps synaptic saturation led to lower semantic distance after an interval of wake, while an interval of sleep allowed the synapses to normalise and participants to think of more distantly related words. According to Tononi and Cirelli, this process of sleep-dependent down-selection of synapses is at the heart of the role of sleep in memory consolidation and integration (Tononi & Cirelli, 2014) – in other words, memory restructuring. On the other hand, performance on the AUT benefitted from time spent awake rather than asleep, and from the specific time of day that the task was completed. These results indicate that there are many aspects of cognitive processing that can affect task performance, and that sleep is only one of the elements that can change our behaviour.

Together, the results presented throughout this thesis indicate that sleep may indeed promote memory restructuring in certain tasks. However, the sleep benefits we found were not always explicit (Chapter 2) or consistent (Chapter 3), and in one instance were entirely absent (AUT in Chapter 4). A previous systematic review found that the specific task chosen by the researcher appeared to have the biggest effect on whether they would find sleep-related benefits or not (Lerner & Gluck, 2019). Successful tasks may thus have particular properties that respond well to sleep, and finding the boundary conditions of memory restructuring during sleep could tell us a lot about its underlying mechanisms.

5.4.2 The roles of wakefulness, REM, and NREM sleep

The experiments in this thesis were conducted with the expectation that sleep generally benefits processes that increase memory restructuring, whether that expresses itself through the extraction of regularities, the construction of remote associations, or improved creativity. However, based on our results, and on the literature, it may be interesting to consider the roles that wakefulness, REM, and NREM sleep play in processes of memory reorganisation.

In section 5.3.3 I explained that some have proposed that the initial selection of relevant memories already takes place during or shortly after encoding – in other words, during wakefulness (Stickgold & Walker, 2013). There have been several experiments which indicate a role for offline processing during wakefulness in memory restructuring. For instance, studies on transitive inference (Ellenbogen et al., 2007), statistical learning (Durrant et al., 2011), decontextualisation of memories (Cox, Tijdens, Meeter, Sweegers, & Talamini, 2014), and memory generalisation (Sweegers & Talamini, 2014), have all found benefits of periods of wakefulness before retesting. These findings fit with the AUT results in Chapter 4, where I found a benefit of an interval containing wake rather than sleep. In contrast, the control group from Experiment 1 in Chapter 3 displayed no particular benefits from wake TMR on remote associations, in line with previous experiments looking at relational memory (Alger & Payne, 2016; Lau et al., 2010). Moreover, while I did not collect a wake TMR group in Chapter 2, a previous experiment using the same task has shown no benefit of wake TMR on sequence-specific skill or explicit sequence memory (Cousins et al., 2014). However, some experiments not employing TMR have found wake-related improvements on the SRTT (Keisler, Ashe, & Willingham, 2007; Robertson et al., 2004). In short, offline processing during wakefulness may indeed benefit memory restructuring in some, but not all, tasks.

REM sleep has been thought to be particularly important for memory restructuring. Several theoretical papers have hypothesised that REM sleep benefits associative processes (Landmann et al., 2014; Lewis et al., 2018). This is corroborated by experiments where participants awoken from REM sleep showed stronger priming effects (Stickgold, Scott, Rittenhouse, & Hobson, 1999) and higher ability to solve anagrams (Walker, Liston, et al., 2002). Another well-known study showed that participants who entered REM sleep during a nap were subsequently more likely to use previously primed words to solve remote associates test (RAT) problems, and this was not associated with improved memory for those words (Cai et al., 2009). Two other studies have found correlations between time spent in REM sleep and increased memory restructuring, as measured by the weather prediction task (Barsky et al., 2015) and relational memory (Alger & Payne, 2016). In this thesis, I only found evidence of REM sleep involvement in Chapter 3, where we employed a task similar to the one used by Alger and Payne (2016). The results of the first experiment in that chapter thus fit nicely with the literature, but as has been pointed out before, the REM TMR-related increases in remote associations did not appear to be reliable. Combined with the fact that there was no indication of REM sleep involvement in memory restructuring in the other experimental chapters, the role of REM sleep in this process remains unclear.

On the other hand, the evidence regarding NREM sleep involvement in the reorganisation of memory is more consistent. Again, theoretical papers have highlighted the role of NREM sleep – and SWS in particular – in the reorganisation of memory. There seems to be somewhat of a consensus that NREM sleep is important for the abstraction of rules and integration of new information into existing schemas (Landmann et al., 2014; Lewis & Durrant, 2011; Lewis et al., 2018; Rasch & Born, 2013). Experiments using the SRTT, including my own in Chapter 2, have consistently shown that NREM sleep and its electrophysiological features are involved in increasing performance on this task (Cousins et al., 2014, 2016; Diekelmann et al., 2016; Wilhelm et al., 2013). Similarly, studies using the number reduction task, where insight into the hidden rule that governs the task greatly improves performance, have found that this insight is promoted by SWS (Verleger, Rose, Wagner, Yordanova, & Kolev, 2013; Yordanova et al., 2008). Other evidence comes from statistical tone learning (Durrant et al., 2013, 2011), artificial grammar learning (Batterink & Paller, 2017; Gaskell et al., 2014), and creative problem solving (Beijamini et al., 2014; Sanders et al., 2019). The results from Chapter 2, where I found behavioural performance increases after TMR in SWS, are completely in line with findings from the literature. Perhaps more interestingly, I also found involvement of NREM sleep in the VGT in Chapter 4. Specifically, the increase in semantic distance between given nouns and generated verbs in the random measure correlated with time spent in Stage 2 sleep. This sleep stage is sometimes overlooked in favour of SWS, but these results indicate that its relationship to memory restructuring deserves more attention.

Overall, wakefulness, REM, and NREM sleep all seem to have some involvement in memory restructuring. Indeed, it seems likely that processes during both wake and sleep work together to support the complex nature of memory and memory restructuring. Future research should determine what these processes are, and whether they are perhaps complementary across different states of consciousness.

5.5 The link between reprocessing and restructuring

The title of this thesis is “Sleep’s role in the reprocessing and restructuring of memory”. I chose my behavioural tasks based on the fact that they were likely to contain some element of memory restructuring that could be sleep-dependent. But precisely how could memory restructuring be accomplished? The central theory that has guided this work is that memory reactivation, or reprocessing more generally, drives restructuring. In fact, in both Chapters 2 and 3, I specifically looked at the link between reprocessing and restructuring, by using TMR. This method explicitly

assumes that memory reactivation is at the core of what drives sleep changes in memory consolidation and restructuring. Insofar as TMR indeed triggers reactivation – and our results from Chapter 2 argue fairly convincingly that it does – this establishes a link between reactivation and behavioural changes.

The idea of memory reactivation being the driving factor behind memory reorganisation plays a key role in many theoretical models, such as Active System Consolidation (Diekelmann & Born, 2010; Rasch & Born, 2013), Information Overlap to Extract (iOtA) (Lewis & Durrant, 2011), Recurrency and Episodic Memory Results in Generalisation (REMERGE) (Kumaran & McClelland, 2012), and Landmann's model which looks at schema formation, integration, and disintegration (Landmann et al., 2014). Admittedly, not everyone agrees that reactivation is the key to sleep's role in memory (Tononi & Cirelli, 2003, 2014). Even among the models that do centre reactivation, the precise mechanisms and underlying neurophysiology leading to memory restructuring are still debated.

In the introduction of this thesis we discussed four well-known general models of sleep and memory: The Dual Process Hypothesis, the Sequential Hypothesis, the Active System Consolidation Hypothesis, and the Synaptic Homeostasis Hypothesis. It is worthwhile to consider which of these models is best supported by the experiments in this thesis.

Clearly, the Dual Process Hypothesis (Plihal & Born, 1997; Wagner et al., 2001), which posits that different types of memory benefit from different sleep stages, does not fit with the evidence we found. The clearest example of this comes from Chapter 2, where we found that a procedural task benefitted from TMR during SWS and not REM sleep, contrary to what the model predicts. Admittedly, this model has gone out of favour, because it does not account for many findings in the literature, including some of the ones presented in this thesis.

In terms of the Sequential Hypothesis (Ambrosini & Giuditta, 2001; Giuditta, 2014; Giuditta et al., 1995), which outlines the importance of the cyclical progression of sleep stages, this appears to be partially supported by our experiments. Chapter 2, for instance, showed behavioural benefits relating to SWS, but electrophysiological differences during REM sleep. As discussed in section 5.3.2, this hypothesis could perhaps explain the lack of successful REM TMR experiments. Nevertheless, the lack of reliable findings in Chapter 3 remain puzzling in light of this model.

The Active System Consolidation Hypothesis (Diekelmann & Born, 2010; Rasch & Born, 2013), then, mainly outlines a role for SWS in the reactivation of memory traces. Again, this fits with the benefits

of SWS TMR we found in Chapter 2. However, Chapter 3 did not show any behavioural improvements relating to TMR during SWS. Within this model, the role of REM sleep is hypothesised to be a stabilising one, supporting the transportation of memory traces to the long-term store through synaptic consolidation. Although we could not test this theory, it fits with the idea that the sleep stages play complementary roles, like in the Sequential Hypothesis. Similarly to that hypothesis, this part of the Active System Consolidation Hypothesis is not contradicted by our data, but not necessarily supported by it either.

Finally, the Synaptic Homeostasis Hypothesis (Tononi & Cirelli, 2003, 2014) places the emphasis on large-scale synaptic downscaling during SWS, which is thought to improve the signal-to-noise ratio in the cortex and lead to memory consolidation of particular memories. Since the role for memory reactivation in this theory is very minor, and since the majority of the experiments in this thesis have used TMR, it is difficult to say much about how well this model explains our findings. Certainly, it is possible that synaptic downscaling worked together with memory reactivation (as triggered by TMR) in for example Chapter 2. However, this was not directly tested in our experiments. The best evidence in favour of this hypothesis comes from Chapter 4, where we found that semantic distance in the VGT became lower (albeit non-significantly) after wake, while sleep increased this back to about baseline level. As discussed, one explanation for this could be that the synaptic saturation which is thought to take place during wakefulness led to a decrease in semantic distance of the words participants came up with after this interval filled with wake. On the other hand, the sleep interval, filled with synaptic downscaling, could have allowed for a 'reset' of VGT performance. However, it remains difficult to incorporate the other findings from Chapter 4. In the AUT, we found that an interval of wakefulness was actually beneficial for performance, whereas sleep did not appear to have any effect. This does not seem to be in line with the idea that increased synaptic strength during wakefulness saturates the ability to learn. It would be interesting to look at the cellular effects of sleep and wakefulness in these two tasks.

There are also models of sleep and memory which focus particularly on memory restructuring, which allows us to speak a bit more specifically about the type of experiments used in this thesis. The most notable example is the broader iOtA model (BiOtA), which has recently outlined explicit mechanisms for the involvement of REM and NREM sleep in the reorganisation of existing knowledge. In this model, NREM sleep is responsible for the consolidation of memories, the abstraction of gist, and the formation of schemas through concurrent replay of related memories. Subsequent REM sleep then allows for the detection of similarities between schemas that initially

were not related, by replay of salient and random schemas during this stage. When similarities are detected, the brain is able to form novel connections between related concepts, eventually leading to a restructuring of semantic knowledge networks.

The NREM part of this model is well-supported by experimental evidence, some of which has been presented in section 5.4.2. I did not make use of a task which allows for gist abstraction, but one could consider our hypothesis that the learned associations in Chapter 3 would benefit from SWS TMR as fitting within the mechanisms of the BiOtA model. In that sense, our results, which showed no effects of SWS TMR, do not provide support for this model. Notably, BiOtA specifically talks about declarative memories (and even more specifically, creativity). Nevertheless, procedural memories also benefit from rule extraction and generalisation, and a motor schema may arise after extended training (Landmann et al., 2014). Our results from Chapter 2 fit with the idea that reactivation during SWS promotes rule extraction also in procedural memories.

The role of REM as outlined in the BiOtA model is somewhat more contentious. The remote associations in Chapter 3 could be considered to test the REM part of BiOtA, since these were associations between related concepts which shared some similarities. While we found some very promising results in the first experiment of that chapter, including indications that the REM TMR effect we found was driven by some kind of memory reprocessing during REM sleep, the subsequent replication experiments showed that these results were not consistent. Overall, this part of BiOtA is thus not supported by the evidence we have found.

There are others that have outlined a slightly different role for REM in the reorganisation of memories during consolidation, for instance through the disintegration of schemas into individual elements, which then allows for the creation of new schemas (Landmann et al., 2015, 2014). Unfortunately, we cannot draw any conclusions about this explanation based on our experiments. Some even argue that this proposed role of REM in restructuring may not be due to memory reactivation. Rather, they say, this may be due to non-specific downgrading of representations during REM, or because of random activation (not reactivation) within neural networks (Dudai et al., 2015; Grosmark, Mizuseki, Pastalkova, Diba, & Buzsáki, 2012). However, our findings that reprocessing does appear to take place during REM somewhat contradict these explanations.

In conclusion, the link between memory reprocessing and restructuring is an explicit part of most models of sleep and memory, although there are some exceptions. Within this thesis, the results from Chapter 2 most strongly demonstrate this link, given that we found evidence of memory

reactivation and implicit memory restructuring effects of TMR. The results from Chapter 3, on the other hand, are inconclusive. Chapter 4 did not include any measurement of reactivation, nor any TMR, but the promising findings show that this might be a viable option for the future.

5.6 Future directions

Throughout this discussion, I have noted some directions for future research. Although we now know that sleep plays an important role in the reprocessing and restructuring of memory, the work presented throughout this thesis has also brought to light a lot of uncertainties and open questions. I want to highlight some of the most important ones here.

First, Chapter 4 has outlined some very interesting possibilities with regards to the role of sleep in creative processes. Creativity has likely always been an important skill, but this may be exceedingly the case right now. In a world full of pandemics, climate change, and economic crises, we need creative solutions to the problems we face. Understanding the processes and mechanisms of creativity, including those that may involve sleep, is thus of paramount importance. In the VGT, we found a positive effect of an interval containing sleep. This effect could be further explored using TMR, and such an experiment would also clarify the role that memory reactivation plays in this task. In the AUT, we did not find any sleep-related benefits, but TMR during sleep may bring these out if they exist and were overshadowed by the time of day effect in our study. As mentioned before, a nap study could also clarify the particular functions of sleep, wake, and time of day here.

This also ties in with the role of REM sleep, which has been said to be important for creativity (Cai et al., 2009; Landmann et al., 2014; Lewis et al., 2018). Throughout this thesis, I have noted that the functions of REM sleep remain particularly unclear, compared with those of NREM sleep and its electrophysiological features. In rodents, there is relatively consistent evidence which shows that REM sleep is involved in memory (thoroughly reviewed in Boyce, Williams, & Adamantidis, 2017; Poe, Walsh, & Bjorness, 2010; Smith, 2011). However, in humans, the overall pattern of studies has been contradictory (Rasch & Born, 2013). It has been suggested that in humans, REM sleep plays a role in procedural but not declarative memory (Smith, 1995, 2001). Our results from Chapter 2 suggest that memory reactivation during REM does not directly influence the SRTT, but REM sleep may still be involved in less obvious ways. Chapter 3 started out with the promising finding of a REM TMR effect on remote associations, but this did not replicate. We also did not find any associations between creative task performance and REM sleep in Chapter 4, although this requires

further investigation. In short, the role of REM sleep in human memory is a mystery that deserves more attention.

In relation to the null-findings in the replication experiments of Chapter 3, the issue of individual differences was brought up. Currently, there is very little work looking at why certain people appear to benefit greatly from sleep and TMR effects on memory, while others do not. Possible participating factors could be sleep characteristics (Kirov et al., 2015), working memory (Fenn & Hambrick, 2012), sex (Diekelmann et al., 2016), age (Gui et al., 2017), and likely many others. Establishing and investigating such factors may also give us insights into the mechanisms of sleep-dependent memory consolidation and restructuring.

A consistent theme throughout this thesis has been the underlying mechanisms of memory restructuring. Many theoretical models include a prominent role for memory reactivation in this process, but among those models there is no consensus on the precise neural mechanisms that accompany memory reactivation (Kumaran & McClelland, 2012; Landmann et al., 2014; Lewis et al., 2018; Rasch & Born, 2013). Moreover, within the field there is no consensus about whether memory reactivation is even the driving factor behind sleep effects on memory (Tononi & Cirelli, 2014), nor about whether memory is indeed restructured or reorganised in all instances where this has been said to be the case (Lerner & Gluck, 2019). In sum, our understanding of the neural mechanisms behind these processes has a long way to go.

5.7 Conclusion

The current thesis investigated the role of sleep in memory reprocessing and restructuring. I showed that memory reactivation takes place during SWS, and that some reprocessing likely also happens during REM sleep. Furthermore, this reprocessing may be selective, and may benefit some memories more than others. With regards to restructuring, it appears that there may also be certain boundary conditions which allow sleep to play a role in this process. While some of the experiments in this thesis benefitted from (TMR during) sleep, the results were not uniform. It is clear that the specific task and features within that task appear to play an important role in whether sleep affects subsequent performance. Moreover, processes during sleep and wakefulness likely work together to support the complex nature of memory and memory restructuring. Finally, I determined that memory reactivation is a probable mechanism of memory restructuring, especially in SWS, though future research must determine how exactly this is accomplished.

References

Abe, M., Schambra, H., Wassermann, E. M., Luckenbaugh, D., Schweighofer, N., & Cohen, L. G. (2011). Reward improves long-term retention of a motor memory through induction of offline memory gains. *Current Biology*, *21*(7), 557–562.

Abraham, A., & Bubic, A. (2015). Semantic memory as the root of imagination. *Frontiers in Psychology*, *6*, 325.

Aeschbach, D., Cutler, A. J., & Ronda, J. M. (2008). A Role for Non-Rapid-Eye-Movement Sleep Homeostasis in Perceptual Learning. *Journal of Neuroscience*, *28*(11), 2766–2772.

Ai, S., Yin, Y., Chen, Y., Wang, C., Sun, Y., Tang, X., … Shi, J. (2018). Promoting subjective preferences in simple economic choices during nap. *ELife*, *7*, e40583.

Åkerstedt, T., & Gillberg, M. (1990). Subjective and Objective Sleepiness in the Active Individual. *International Journal of Neuroscience*, *52*(1–2), 29–37.

Albouy, G., Sterpenich, V., Balteau, E., Vandewalle, G., Desseilles, M., Dang-Vu, T., … Maquet, P. (2008). Both the Hippocampus and Striatum Are Involved in Consolidation of Motor Sequence Memory. *Neuron*, *58*(2), 261–272.

Alger, S. E., & Payne, J. D. (2016). The differential effects of emotional salience on direct associative and relational memory during a nap. *Cognitive, Affective and Behavioral Neuroscience*, *16*, 1150–1163.

Ambrosini, M. V., & Giuditta. (2001). Learning and sleep: The sequential hypothesis. *Sleep Medicine Reviews*, *5*(6), 477–490.

Anderson, J. R., & Ross, B. H. (1980). Evidence against a semantic-episodic distinction. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(5), 441–466.

Andrillon, T., Pressnitzer, D., Léger, D., & Kouider, S. (2017). Formation and suppression of acoustic memories during human sleep. *Nature Communications*, *8*, 179.

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.

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.e4.

Arden, R., Chavez, R. S., Grazioplene, R., & Jung, R. E. (2010). Neuroimaging creativity: A

psychometric view. *Behavioural Brain Research*, *214*(2), 143–156.

Arnal, P. J., Thorey, V., Debellemaniere, E., Ballard, M. E., Bou Hernandez, A., Guillot, A., … Sauvet, F. (2020). The Dreem Headband compared to Polysomnography for EEG Signal Acquisition and Sleep Staging. *Sleep*, zsaa097.

Arzi, A., Shedlesky, L., Ben-Shaul, M., Nasser, K., Oksenberg, A., Hairston, I. S., & Sobel, N. (2012). Humans can learn new information during sleep. *Nature Neuroscience*, *15*(10), 1460–1465.

Astori, S., Wimmer, R. D., & Lüthi, A. (2013). Manipulating sleep spindles - expanding views on sleep, memory, and disease. *Trends in Neurosciences*, *36*(12), 738–748.

Baas, M., De Dreu, C. K. W., & Nijstad, B. A. (2008). A Meta-Analysis of 25 Years of Mood-Creativity Research: Hedonic Tone, Activation, or Regulatory Focus? *Psychological Bulletin*, *134*(6), 779–806.

Baird, B., Smallwood, J., Mrazek, M. D., Kam, J. W. Y., Franklin, M. S., & Schooler, J. W. (2012). Inspired by Distraction: Mind Wandering Facilitates Creative Incubation. *Psychological Science*, *23*(10), 1117–1122.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*, 379–384.

Barakat, M., Doyon, J., Debas, K., Vandewalle, G., Morin, A., Poirier, G., … Carrier, J. (2011). Fast and slow spindle involvement in the consolidation of a new motor sequence. *Behavioural Brain Research*, *217*(1), 117–121.

Barrett, T. R., & Ekstrand, B. R. (1972). Effect of sleep on memory: III. Controlling for time-of-day effects. *Journal of Experimental Psychology*, *96*(2), 321–327.

Barsky, M. M., Tucker, M. A., & Stickgold, R. (2015). REM sleep enhancement of probabilistic classification learning is sensitive to subsequent interference. *Neurobiology of Learning and Memory*, *122*, 63–68.

Basner, M., & Dinges, D. F. (2011). Maximizing Sensitivity of the Psychomotor Vigilance Test (PVT) to Sleep Loss. *Sleep*, *34*(5), 581–591.

Bastuji, H., & García-Larrea, L. (1999). Evoked potentials as a tool for the investigation of human sleep. *Sleep Medicine Reviews*, *3*(1), 23–45.

Battaglia, F. P., Sutherland, G. R., Cowen, S. L., Mc Naughton, B. L., & Harris, K. D. (2005). Firing rate modulation: A simple statistical view of memory trace reactivation. *Neural Networks*, *18*(9), 1280–1291.

Batterink, L. J., & Paller, K. A. (2017). Sleep-based memory processing facilitates grammatical

generalization: Evidence from targeted memory reactivation. *Brain and Language*, *167*, 83–93.

Beaty, R. E., Benedek, M., Silvia, P. J., & Schacter, D. L. (2016). Creative Cognition and Brain Network Dynamics. *Trends in Cognitive Sciences*, *20*(2), 87–95.

Beaty, R. E., & Silvia, P. J. (2012). Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(4), 309–319.

Beijamini, F., Pereira, S. I. R., Cini, F. A., & Louzada, F. M. (2014). After being challenged by a video game problem, sleep increases the chance to solve it. *PLoS ONE*, *9*(1), e84342.

Belal, S., Cousins, J. N., El-Deredy, W., Parkes, L., Schneider, J., Tsujimura, H., … Lewis, P. (2018). Identification of memory reactivation during sleep by EEG classification. *NeuroImage*, *176*, 203–214.

Bendor, D., & Wilson, M. A. (2012). Biasing the content of hippocampal replay during sleep. *Nature Neuroscience*, *15*(10), 1439–1444.

Benedek, M., Kenett, Y. N., Umdasch, K., Anaki, D., Faust, M., & Neubauer, A. C. (2017). How semantic memory structure and intelligence contribute to creative thought: A network science approach. *Thinking & Reasoning*, *23*(2), 158–183.

Benedek, M., Könen, T., & Neubauer, A. C. (2012). Associative abilities underlying creativity. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(3), 273–281.

Benedek, M., & Neubauer, A. C. (2013). Revisiting Mednick's Model on Creativity-Related Differences in Associative Hierarchies. Evidence for a Common Path to Uncommon Thought. *The Journal of Creative Behavior*, *47*(4), 273–289.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300.

Bergmann, T. O., Mölle, M., Diedrichs, J., Born, J., & Siebner, H. R. (2012). Sleep spindle-related reactivation of category-specific cortical regions after learning face-scene associations. *NeuroImage*, *59*(3), 2733–2742.

Bernard, M., Kenett, Y., Ovando-Tellez, M., Benedek, M., & Volle, E. (2019). Building Individual Semantic Networks and Exploring their Relationships with Creativity. In *The 41st Annual Meeting of the Cognitive Science Society, At Montreal, Canada* (pp. 138–144).

Berry, R. B., Brooks, R., Gamaldo, C. E., Harding, S. M., Lloyd, R. M., Marcus, C. L., & Vaughn, B. V. (2015). *The AASM Manual for the Scoring of Sleep and Associated Events. Rules, terminology and technical specifications.* Darien, IL: American Academy of Sleep Medicine.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527–536.

Booth, V., & Poe, G. R. (2006). Input source and strength influences overall firing phase of model hippocampal CA1 pyramidal cells during theta: Relevance to REM sleep reactivation and memory consolidation. *Hippocampus*, *16*(2), 161–173.

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

Boyce, R., Williams, S., & Adamantidis, A. (2017). REM sleep and memory. *Current Opinion in Neurobiology*, *44*, 167–177.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Brodt, S., Pöhlchen, D., Täumer, E., Gais, S., & Schönauer, M. (2018). Incubation, not sleep, aids problem-solving. *Sleep*, *41*(10).

Browne, B. A., & Cruse, D. F. (1988). The Incubation Effect: Illusion or Illumination? *Human Performance*, *1*(3), 177–185.

Brualla, J., Romero, M. F., Serrano, M., & Valdizán, J. R. (1998). Auditory event-related potentials to semantic priming during sleep. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *108*(3), 283–290.

Brysbaert, M. (2019). How Many Participants Do We Have to Include in Properly Powered Experiments? A Tutorial of Power Analysis with Reference Tables. *Journal of Cognition*, *2*(1), 16.

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition*, *1*(1), 9.

Buckelmüller, J., Landolt, H. P., Stassen, H. H., & Achermann, P. (2006). Trait-like individual differences in the human sleep electroencephalogram. *Neuroscience*, *138*(1), 351–356.

Bunsey, M., & Eichenbaum, H. (1996). Conservation of hippocampal memory function in rats and humans. *Nature*, *379*, 255–257.

Burgess, N., Donnett, J. G., & O'Keefe, J. (1998). The Representation of Space and the Hippocampus in Rats, Robots and Humans. *Zeitschrift Für Naturforschung C*, *53*(7–8), 504–509.

Busby, K. A., Mercier, L., & Pivik, R. T. (1994). Ontogenetic variations in auditory arousal threshold during sleep. *Psychophysiology*, *31*(2), 182–188.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M.

R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*, 365–376.

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index (PSQI): A new instrument for psychiatric research and practice. *Psychiatry Research*, *28*, 193–213.

Buzsáki, G. (2015). Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. *Hippocampus*, *25*(10), 1073–1188.

Cai, D. J., 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 of the United States of America*, *106*(25), 10130–10134.

Cairney, S. A., Durrant, S. J., Hulleman, J., & Lewis, P. A. (2014). Targeted Memory Reactivation During Slow Wave Sleep Facilitates Emotional Memory Consolidation. *Sleep*, *37*(4), 701–707.

Cairney, S. A., Guttesen, A. á. V., El Marj, N., & Staresina, B. P. (2018). Memory Consolidation Is Linked to Spindle-Mediated Information Processing during Sleep. *Current Biology*, *28*(6), 948-954.e4.

Cairney, S. A., Lindsay, S., Sobczak, J. M., Paller, K. A., & Gaskell, M. G. (2016). The Benefits of Targeted Memory Reactivation for Consolidation in Sleep are Contingent on Memory Accuracy and Direct Cue-Memory Associations. *Sleep*, *39*(5), 1139–1150.

Cairney, S. A., Sobczak, J. M., Lindsay, S., & Gaskell, M. G. (2017). Mechanisms of Memory Retrieval in Slow-Wave Sleep. *Sleep*, *40*(9), zsx114.

Carr, M. F., Jadhav, S. P., & Frank, L. M. (2011). Hippocampal replay in the awake state: A potential substrate for memory consolidation and retrieval. *Nature Neuroscience*, *14*, 147–153.

Cash, S. S., Halgren, E., Dehghani, N., Rossetti, A. O., Thesen, T., Wang, C. M., … Ulbert, I. (2009). The human K-complex represents an isolated cortical down-state. *Science*, *324*(5930), 1084–1087.

Cellini, N., & Cappuzo, A. (2018). Shaping memory consolidation via targeted memory. *Annals of the New York Academy of Sciences*, *1426*, 52–71.

Clemens, Z., Fabó, D., & Halász, P. (2005). Overnight verbal memory retention correlates with the number of sleep spindles. *Neuroscience*, *132*(2), 529–535.

Clemens, Z., Fabó, D., & Halász, P. (2006). Twenty-four hours retention of visuospatial memory correlates with the number of parietal sleep spindles. *Neuroscience Letters*, *403*(1–2), 52–56.

Cohen, D. A., Pascual-Leone, A., Press, D. Z., & Robertson, E. M. (2005). Off-line learning of motor

skill memory: A double dissociation of goal and movement. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(50), 18237–18241.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, *65*(3), 145–153.

Cohen, M. X. (2014). *Analyzing neural time series data: Theory and practice* (1st ed.). MIT Press.

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, *210*(4466), 207–210.

Colzato, L. S., Ozturk, A., & Hommel, B. (2012). Meditate to Create: The Impact of Focused-Attention and Open-Monitoring Training on Convergent and Divergent Thinking. *Frontiers in Psychology*, *3*, 116.

Cordi, M. J., Diekelmann, S., Born, J., & Rasch, B. (2014). No effect of odor-induced memory reactivation during REM sleep on declarative memory stability. *Frontiers in Systems Neuroscience*, *8*, 157.

Cousins, J. N., El-Deredy, W., Parkes, L. M., Hennies, N., & Lewis, P. A. (2014). Cued Memory Reactivation during Slow-Wave Sleep Promotes Explicit Knowledge of a Motor Sequence. *Journal of Neuroscience*, *34*(48), 15870–15876.

Cousins, J. N., El-Deredy, W., Parkes, L. M., Hennies, N., & Lewis, P. A. (2016). Cued Reactivation of Motor Learning during Sleep Leads to Overnight Changes in Functional Brain Activity and Connectivity. *PLOS Biology*, *14*(5), e1002451.

Cox, R., Tijdens, R. R., Meeter, M. M., Sweegers, C. C. G., & Talamini, L. M. (2014). Time, not sleep, unbinds contexts from item memory. *PloS One*, *9*(2), e88307.

Creery, J. D., Oudiette, D., Antony, J. W., & Paller, K. A. (2015). Targeted Memory Reactivation during Sleep Depends on Prior Learning. *Sleep*, *38*(5), 755–763.

Cropley, A. (2006). In Praise of Convergent Thinking. *Creativity Research Journal*, *18*(3), 391–404.

Datta, S., & Hobson, J. A. (2000). The rat as an experimental model for sleep neurophysiology. *Behavioral Neuroscience*, *114*(6), 1239–1244.

Datta, S., & MacLean, R. R. (2007). Neurobiological mechanisms for the regulation of mammalian sleep-wake behavior: Reinterpretation of historical evidence and inclusion of contemporary cellular and molecular evidence. *Neuroscience and Biobehavioral Reviews*, *31*(5), 775–824.

Davis, M. A. (2009). Understanding the relationship between mood and creativity: A meta-analysis. *Organizational Behavior and Human Decision Processes*, *108*(1), 25–38.

De Gennaro, L., & Ferrara, M. (2003). Sleep spindles: An overview. *Sleep Medicine Reviews*, *7*(5), 423–440.

De Lavilléon, G., Lacroix, M. M., Rondi-Reig, L., & Benchenane, K. (2015). Explicit memory creation during sleep demonstrates a causal role of place cells in navigation. *Nature Neuroscience*, *18*(4), 493–495.

Debarnot, U., Creveaux, T., Collet, C., Doyon, J., & Guillot, A. (2009). Sleep Contribution to Motor Memory Consolidation: A Motor Imagery Study. *Sleep*, *32*(12), 1559–1565.

Debellemaniere, E., Chambon, S., Pinaud, C., Thorey, V., Dehaene, D., Léger, D., … Galtier, M. N. (2018). Performance of an Ambulatory Dry-EEG Device for Auditory Closed-Loop Stimulation of Sleep Slow Oscillations in the Home Environment. *Frontiers in Human Neuroscience*, *12*, 88.

Dehaene, S., & King, J.-R. (2016). Decoding the Dynamics of Conscious Perception: The Temporal Generalization Method. In G. Buzsáki & Y. Christen (Eds.), *Micro-, Meso- and Macro-Dynamics of the Brain. Research and Perspectives in Neurosciences* (pp. 85–97). Springer.

Delacre, M., Lakens, D., & Leys, C. (2017). Why Psychologists Should by Default Use Welch's t-test Instead of Student's t-test. *International Review of Social Psychology*, *30*(1), 92–101.

Deuker, L., Olligs, J., Fell, J., Kranz, T. A., Mormann, F., Montag, C., … Axmacher, N. (2013). Memory Consolidation by Replay of Stimulus-Specific Neural Activity. *Journal of Neuroscience*, *33*(49), 19373–19383.

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, *11*(2), 114–126.

Diekelmann, S., Born, J., & Rasch, B. (2016). Increasing Explicit Sequence Knowledge by Odor Cueing during Sleep in Men but not Women. *Frontiers in Behavioral Neuroscience*, *10*, 74.

Diekelmann, S., Born, J., & Wagner, U. (2010). Sleep enhances false memories depending on general memory performance. *Behavioural Brain Research*, *208*(2), 425–429.

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

Dijksterhuis, A., & Meurs, T. (2006). Where creativity resides: The generative power of unconscious thought. *Consciousness and Cognition*, *15*(1), 135–146.

Dijksterhuis, A., & Nordgren, L. F. (2006). A Theory of Unconscious Thought. *Perspectives on Psychological Science*, *1*(2), 95–109.

Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, &*

Computers, *17*(6), 652–655.

Djonlagic, I., Rosenfeld, A., Shohamy, D., Myers, C., Gluck, M., & Stickgold, R. (2009). Sleep enhances category learning. *Learning and Memory*, *16*, 751–755.

Doran, S. M., Wessel, T., Kilduff, T. S., Turek, F., & Renger, J. J. (2008). Translational models of sleep and sleep disorders. In R. A. McArthur & F. Borsini (Eds.), *Animal and Translational Models for CNS Drug Discovery* (Vol. 1, pp. 395–456). Academic Press.

Doyon, J., & Benali, H. (2005). Reorganization and plasticity in the adult brain during learning of motor skills. *Current Opinion in Neurobiology*, *15*(2), 161–167.

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

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576–585.

Dudai, Y., Karni, A., & Born, J. (2015). The Consolidation and Transformation of Memory. *Neuron*, *88*(1), 20–32.

Dupret, D., O'Neill, J., Pleydell-Bouverie, B., & Csicsvari, J. (2010). The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nature Neuroscience*, *13*, 995–1002.

Durrant, S. J., Cairney, S. A., & Lewis, P. A. (2013). Overnight Consolidation Aids the Transfer of Statistical Knowledge from the Medial Temporal Lobe to the Striatum. *Cerebral Cortex*, *23*(10), 2467–2478.

Durrant, S. J., Cairney, S. A., & Lewis, P. A. (2016). Cross-modal transfer of statistical information benefits from sleep. *Cortex*, *78*, 85–99.

Durrant, S. J., Taylor, C., Cairney, S., & Lewis, P. A. (2011). Sleep-dependent consolidation of statistical learning. *Neuropsychologia*, *49*(5), 1322–1331.

Eichenbaum, H. (2004). Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron*, *44*(1), 109–120.

Ellenbogen, J. M., Hu, P. T., Payne, J. D., Titone, D., & Walker, M. P. (2007). Human relational memory requires time and sleep. *Proceedings of the National Academy of Sciences*, *104*(18), 7723–7728.

Euston, D. R., Tatsuno, M., & McNaughton, B. L. (2007). Fast-forward playback of recent memory

sequences in prefrontal cortex during sleep. *Science*, *318*(5853), 1147–1150.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160.

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

Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature*, *425*, 614–616.

Fernandez, L. M. J., & Lüthi, A. (2020). Sleep Spindles: Mechanisms and Functions. *Physiological Reviews*, *100*(2), 805–868.

Ficca, G., Lombardo, P., Rossi, L., & Salzarulo, P. (2000). Morning recall of verbal material depends on prior sleep organization. *Behavioural Brain Research*, *112*(1–2), 159–163.

Ficca, G., & Salzarulo, P. (2004). What is sleep is for memory. *Sleep Medicine*, *5*(3), 225–230.

Fischer, S., & Born, J. (2009). Anticipated Reward Enhances Offline Learning During Sleep. *Journal of Experimental Psychology: Learning Memory and Cognition*, *35*(6), 1586–1593.

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.

Fogel, S. M., Smith, C. T., & Cote, K. A. (2007). Dissociable learning-dependent changes in REM and non-REM sleep in declarative and procedural memory systems. *Behavioural Brain Research*, *180*(1), 48–61.

Fowler, M. J., Sullivan, M. J., & Ekstrand, B. R. (1973). Sleep and memory. *Science*, *179*(4070), 302–304.

Fraley, R. C., & Vazire, S. (2014). The N-Pact Factor: Evaluating the Quality of Empirical Journals with Respect to Sample Size and Statistical Power. *PLoS ONE*, *9*(10), e109019.

Friedrich, M., Mölle, M., Friederici, A. D., & Born, J. (2019). The reciprocal relation between sleep and memory in infancy: Memory-dependent adjustment of sleep spindles and spindle-dependent improvement of memories. *Developmental Science*, *22*(2), e12743.

Friedrich, M., Wilhelm, I., Born, J., & Friederici, A. D. (2015). Generalization of word meanings during infant sleep. *Nature Communications*, *6*, 6004.

Friedrich, M., Wilhelm, I., Mölle, M., Born, J., & Friederici, A. D. (2017). The Sleeping Infant Brain

Anticipates Development. *Current Biology*, *27*(15), 2374-2380.e3.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2–18.

Fuentemilla, L., Miró, J., Ripollés, P., Vilà-Balló, A., Juncadella, M., Castañer, S., … Rodríguez-Fornells, A. (2013). Hippocampus-Dependent Strengthening of Targeted Memories via Reactivation during Sleep in Humans. *Current Biology*, *23*(18), 1769–1775.

Gais, S., Mölle, M., Helms, K., & Born, J. (2002). Learning-dependent increases in sleep spindle density. *Journal of Neuroscience*, *22*(15), 6830–6834.

Gais, S., Plihal, W., Wagner, U., & Born, J. (2000). Early sleep triggers memory for early visual discrimination skills. *Nature Neuroscience*, *3*(12), 1335–1339.

Gaskell, M. G., Warker, J., Lindsay, S., Frost, R., Guest, J., Snowdon, R., & Stackhouse, A. (2014). Sleep Underpins the Plasticity of Language Production. *Psychological Science*, *25*(7), 1457–1465.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*(1), 1–38.

Gilhooly, K. J., Fioratou, E., Anthony, S. H., & Wynn, V. (2007). Divergent thinking: Strategies and executive involvement in generating novel uses for familiar objects. *British Journal of Psychology*, *98*(4), 611–625.

Gilhooly, K. J., Georgiou, G. J., & Devery, U. (2013). Incubation and creativity: Do something different. *Thinking & Reasoning*, *19*(2), 137–149.

Gilhooly, K. J., Georgiou, G. J., Garrison, J., Reston, J. D., & Sirota, M. (2012). Don't wait to incubate: Immediate versus delayed incubation in divergent thinking. *Memory and Cognition*, *40*(6), 966–975.

Gilhooly, K. J., Georgiou, G. J., Sirota, M., & Paphiti-Galeano, A. (2015). Incubation and suppression processes in creative problem solving. *Thinking & Reasoning*, *21*(1), 130–146.

Girardeau, G., Benchenane, K., Wiener, S. I., Buzsáki, G., & Zugaro, M. B. (2009). Selective suppression of hippocampal ripples impairs spatial memory. *Nature Neuroscience*, *12*, 1222–1223.

Girardeau, G., Cei, A., & Zugaro, M. (2014). Learning-Induced Plasticity Regulates Hippocampal Sharp Wave-Ripple Drive. *Journal of Neuroscience*, *34*(15), 5176–5183.

Girardeau, G., & Zugaro, M. (2011). Hippocampal ripples and memory consolidation. *Current Opinion in Neurobiology*, *21*(3), 452–459.

Giuditta, A. (2014). Sleep memory processing: the sequential hypothesis. *Frontiers in Systems Neuroscience*, *8*, 219.

Giuditta, A., Ambrosini, M. V., Montagnese, P., Mandile, P., Cotugno, M., Zucconi, G. G., & Vescia, S. (1995). The sequential hypothesis of the function of sleep. *Behavioural Brain Research*, *69*(1–2), 157–166.

Golden, S. S., Ishiura, M., Johnson, C. H., & Kondo, T. (1997). Cyanobacterial circadian rhythms. *Annual Review of Plant Physiology and Plant Molecular Biology*, *48*, 327–354.

Göldi, M., van Poppel, E. A. M., Rasch, B., & Schreiner, T. (2019). Increased neuronal signatures of targeted memory reactivation during slow-wave up states. *Scientific Reports*, *9*, 2715.

Gómez, R. L., Bootzin, R. R., & Nadel, L. (2006). Naps Promote Abstraction in Language-Learning Infants. *Psychological Science*, *17*(8), 670–674.

Görtelmeyer, R. (1985). On the development of a standardized sleep inventory for the assessment of sleep. In S. Kubicki & W. M. Herrmann (Eds.), *Methods of Sleep Research* (pp. 93–98). Stuttgart: Gustav Fischer.

Graveline, Y. M., & Wamsley, E. J. (2017). The impact of sleep on novel concept learning. *Neurobiology of Learning and Memory*, *141*, 19–26.

Gridchyn, I., Schoenenberger, P., O'Neill, J., & Csicsvari, J. (2020). Assembly-Specific Disruption of Hippocampal Replay Leads to Selective Memory Deficit. *Neuron*, *106*(2), 291-300.e6.

Grosmark, A. D., Mizuseki, K., Pastalkova, E., Diba, K., & Buzsáki, G. (2012). REM Sleep Reorganizes Hippocampal Excitability. *Neuron*, *75*(6), 1001–1007.

Guerrien, A., Dujardin, K., Mandai, O., Sockeel, P., & Leconte, P. (1989). Enhancement of memory by auditory stimulation during postlearning REM sleep in humans. *Physiology & Behavior*, *45*(5), 947–950.

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

Guilford, J. (1967). *The nature of human intelligence.* New York: McGraw-Hill.

Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun - An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods*, *47*, 930–944.

Gut, M., Urbanik, A., Forsberg, L., Binder, M., Rymarczyk, K., Sobiecka, B., … Grabowska, A. (2007). Brain correlates of right-handedness. *Acta Neurobiologiae Experimentalis*, *67*(1), 43–51.

Haba-Rubio, J., & Krieger, J. (2012). Evaluation Instruments for Sleep Disorders: A Brief History of

Polysomnography and Sleep Medicine. In R. P.-Y. Chiang & S.-C. J. Kang (Eds.), *Introduction to Modern Sleep Technology* (pp. 19–31). Dordrecht, the Netherlands: Springer Netherlands.

Hao, N., Ku, Y., Liu, M., Hu, Y., Bodner, M., Grabner, R. H., & Fink, A. (2016). Reflection enhances creativity: Beneficial effects of idea evaluation on idea generation. *Brain and Cognition*, *103*, 30–37.

Hao, N., Ku, Y., Liu, M., Hu, Y., Grabner, R. H., & Fink, A. (2014). Enhancing Verbal Creativity via Brief Interventions During an Incubation Interval. *Creativity Research Journal*, *26*(1), 30–38.

Hao, N., Yuan, H., Cheng, R., Wang, Q., & Runco, M. A. (2015). Interaction effect of response medium and working memory capacity on creative idea generation. *Frontiers in Psychology*, *6*, 1582.

Hars, B., Hennevin, E., & Pasques, P. (1985). Improvement of learning by cueing during postlearning paradoxical sleep. *Behavioural Brain Research*, *18*(3), 241–250.

Hauner, K. K., Howard, J. D., Zelano, C., & Gottfried, J. A. (2013). Stimulus-specific enhancement of fear extinction during slow-wave sleep. *Nature Neuroscience*, *16*(11), 1553–1555.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2430.

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, *92*(2), 490–499.

Heine, R. (1914). *Über Wiedererkennen und rückwirkende Hemmung*. Leipzig, Germany: Barth.

Heinen, D. J. P., & Johnson, D. R. (2018). Semantic distance: An automated measure of creativity that is novel and appropriate. *Psychology of Aesthetics, Creativity, and the Arts*, *12*(2), 144–156.

Himmer, L., Müller, E., Gais, S., & Schönauer, M. (2017). Sleep-mediated memory consolidation depends on the level of integration at encoding. *Neurobiology of Learning and Memory*, *137*, 101–106.

Hirase, H., Leinekugel, X., Czurkó, A., Csicsvari, J., & Buzsáki, G. (2001). Firing rates of hippocampal neurons are preserved during subsequent sleep episodes and modified by novel awake experience. *Proceedings of the National Academy of Sciences*, *98*(16), 9386–9390.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of sleepiness: a new approach. *Psychophysiology*, *10*(4), 431–436.

Hołda, M., Głodek, A., Dankiewicz-Berger, M., Skrzypińska, D., & Szmigielska, B. (2020). Ill-Defined

Problem Solving Does Not Benefit From Daytime Napping. *Frontiers in Psychology*, *11*, 559.

Horne, J. A. (1988). Sleep Loss and "Divergent" Thinking Ability. *Sleep*, *11*(6), 528–536.

Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology*, *4*, 97–110.

Howe, T., Wilson, M. A., Ji, D., & Jones, M. W. (2019). Extending evidence for REM-associated replay in hippocampal CA1 place cells. In *Program No. 333.10.2019 Neuroscience Meeting Planner* (p. Online). Chicago, IL: Society for Neuroscience. Retrieved from https://www.abstractsonline.com/pp8/#!/7883/presentation/67466

Hu, P., Stylos-Allan, M., & Walker, M. P. (2006). Sleep facilitates consolidation of emotional declarative memory. *Psychological Science*, *17*(10), 891–898.

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

Huber, R., Felice Ghilardi, M., Massimini, M., & Tononi, G. (2004). Local sleep and learning. *Nature*, *430*(6995), 78–81.

Hutchison, I. C., & Rathore, S. (2015). The role of REM sleep theta activity in emotional memory. *Frontiers in Psychology*, *6*, 1439.

Intons-Peterson, M. J., Rocchi, P., West, T., McLellan, K., & Hackney, A. (1998). Aging, optimal testing times, and negative priming. *Journal of Experimental Psychology: Learning Memory and Cognition*, *24*(2), 362–376.

Irish, M., & Piguet, O. (2013). The pivotal role of semantic memory in remembering the past and imagining the future. *Frontiers in Behavioral Neuroscience*.

JASP Team. (2019). JASP (Version 0.10.2).

Jegou, A., Schabus, M., Gosseries, O., Dahmen, B., Albouy, G., Desseilles, M., … Dang-Vu, T. T. (2019). Cortical reactivations during sleep spindles following declarative learning. *NeuroImage*, *195*, 104–112.

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

Johnson, B. P., Scharf, S. M., Verceles, A. C., & Westlake, K. P. (2019). Use of targeted memory reactivation enhances skill performance during a nap and enhances declarative memory during wake in healthy young adults. *Journal of Sleep Research*, *28*(5), e12832.

Johnson, B. P., Scharf, S. M., & Westlake, K. P. (2018). Targeted Memory Reactivation During Sleep, But Not Wake, Enhances Sensorimotor Skill Performance: A Pilot Study. *Journal of Motor Behavior*, *50*(2), 202–209.

Johnson, L. A., Euston, D. R., Tatsuno, M., & McNaughton, B. L. (2010). Stored-trace reactivation in rat prefrontal cortex is correlated with down-to-up state fluctuation density. *Journal of Neuroscience*, *30*(7), 2650–2661.

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the n-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 615–622.

Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J. M., & Sagi, D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. *Science*, *265*(5172), 679–682.

Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). *Essentials of creativity assessment*. Hoboken, NJ: John Wiley & Sons, Inc.

Keisler, A., Ashe, J., & Willingham, D. T. (2007). Time of day accounts for overnight improvement in sequence learning. *Learning and Memory*, *14*, 669–672.

Kenett, Y. N. (2019). What can quantitative measures of semantic distance tell us about creativity? *Current Opinion in Behavioral Sciences*, *27*, 11–16.

Kenett, Y. N., Anaki, D., & Faust, M. (2014). Investigating the structure of semantic networks in low and high creative persons. *Frontiers in Human Neuroscience*, *8*, 407.

Kenett, Y. N., & Faust, M. (2019). A Semantic Network Cartography of the Creative Mind. *Trends in Cognitive Sciences*, *23*(4), 271–274.

Kim, S. G., Ashe, J., Hendrich, K., Ellermann, J. M., Merkle, H., Uğurbil, K., & Georgopoulos, A. P. (1993). Functional magnetic resonance imaging of motor cortex: Hemispheric asymmetry and handedness. *Science*, *261*(5121), 615–617.

Kirov, R., Kolev, V., Verleger, R., & Yordanova, J. (2015). Labile sleep promotes awareness of abstract knowledge in a serial reaction time task. *Frontiers in Psychology*, *6*, 1354.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., … Nosek, B. A. (2014). Investigating Variation in Replicability. *Social Psychology*, *45*(3), 142–152.

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., … Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Samples and Settings. *Advances in Methods and Practices in Psychological Science*, *1*(4), 443–490.

Kleiner, M., Brainard, D. H., & Pelli, D. (2007). What's new in psychtoolbox-3? In *Perception* (Vol.

36). ECVP Abstract Supplement.

Kohn, N., & Smith, S. M. (2009). Partly versus Completely Out of Your Mind: Effects of Incubation and Distraction on Resolving Fixation. *The Journal of Creative Behavior*, *43*(2), 102–118.

Konkel, A., & Cohen, N. J. (2009). Relational memory and the hippocampus: Representations and methods. *Frontiers in Neuroscience*, *3*(2), 166–174.

Koppel, R. H., & Storm, B. C. (2014). Escaping mental fixation: Incubation and inhibition in creative problem solving. *Memory*, *22*(4), 340–348.

Korman, M., Doyon, J., Doljansky, J., Carrier, J., Dagan, Y., & Karni, A. (2007). Daytime sleep condenses the time course of motor memory consolidation. *Nature Neuroscience*, *10*(9), 1206–1213.

Korman, M., Raz, N., Flash, T., & Karni, A. (2003). Multiple shifts in the representation of a motor sequence during the acquisition of skilled performance. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(21), 12492–12497.

Kriegeskorte, N. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4.

Kudrimoti, H. S., Barnes, C. A., & McNaughton, B. L. (1999). Reactivation of hippocampal cell assemblies: Effects of behavioral state, experience, and EEG dynamics. *Journal of Neuroscience*, *19*(10), 4090–4101.

Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. *Psychological Review*, *119*(3), 573–616.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470.

Lacaux, C., Izabelle, C., Santantonio, G., De Villèle, L., Frain, J., Lubart, T., … Oudiette, D. (2019). Increased creative thinking in narcolepsy. *Brain*, *142*(7), 1988–1999.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2–3), 259–284.

Landmann, N., Kuhn, M., Maier, J.-G., Spiegelhalder, K., Baglioni, C., Frase, L., … Nissen, C. (2015). REM sleep and memory reorganization: Potential relevance for psychiatry and psychotherapy. *Neurobiology of Learning and Memory*, *122*, 28–40.

Landmann, N., Kuhn, M., Piosczyk, H., Feige, B., Baglioni, C., Spiegelhalder, K., … Nissen, C. (2014). The reorganisation of memory during sleep. *Sleep Medicine Reviews*, *18*(6), 531–541.

Lansink, C. S., Goltstein, P. M., Lankelma, J. V., Joosten, R. N. J. M. A., McNaughton, B. L., & Pennartz, C. M. A. (2008). Preferential reactivation of motivationally relevant information in the ventral striatum. *Journal of Neuroscience*, *28*(25), 6372–6382.

Lansink, C. S., Goltstein, P. M., Lankelma, J. V., McNaughton, B. L., & Pennartz, C. M. A. (2009). Hippocampus leads ventral striatum in replay of place-reward information. *PLoS Biology*, *7*(8), e1000173.

Lau, H., Tucker, M. A., & Fishbein, W. (2010). Daytime napping: Effects on human direct associative and relational memory. *Neurobiology of Learning and Memory*, *93*(4), 554–560.

Laventure, S., Fogel, S., Lungu, O., Albouy, G., Sévigny-Dupont, P., Vien, C., … Doyon, J. (2016). NREM2 and Sleep Spindles Are Instrumental to the Consolidation of Motor Sequence Memories. *PLoS Biology*, *14*(3), 1–27.

Laventure, S., Pinsard, B., Lungu, O., Carrier, J., Fogel, S., Benali, H., … Doyon, J. (2018). Beyond spindles: interactions between sleep spindles and boundary frequencies during cued reactivation of motor memory representations. *Sleep*, *41*(9), zsy142.

Lee, A. K., & Wilson, M. A. (2002). Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron*, *36*(6), 1183–1194.

Lehmann, M., Schreiner, T., Seifritz, E., & Rasch, B. (2016). Emotional arousal modulates oscillatory correlates of targeted memory reactivation during NREM, but not REM sleep. *Scientific Reports*, *6*, 39229.

Lerner, I., & Gluck, M. A. (2019). Sleep and the extraction of hidden regularities: A systematic review and the importance of temporal rules. *Sleep Medicine Reviews*, *47*, 39–50.

Lewis, P. A., & Bendor, D. (2019). How Targeted Memory Reactivation Promotes the Selective Strengthening of Memories in Sleep. *Current Biology*, *29*(18), R906–R912.

Lewis, P. A., & Durrant, S. J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends in Cognitive Sciences*, *15*(8), 343–351.

Lewis, P. A., Knoblich, G., & Poe, G. (2018). How Memory Replay in Sleep Boosts Creative Problem-Solving. *Trends in Cognitive Sciences*, *22*(6), 491–503.

Louie, K., & Wilson, M. A. (2001). Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron*, *29*(1), 145–156.

Maddox, W. T., Glass, B. D., Zeithamova, D., Savarie, Z. R., Bowen, C., Matthews, M. D., & Schnyer,

D. M. (2011). The Effects of Sleep Deprivation on Dissociable Prototype Learning Systems. *Sleep*, *34*(3), 253–260.

Maier, N. R. F. (1931). Reasoning and learning. *Psychological Review*, *38*(4), 332–346.

Maingret, N., Girardeau, G., Todorova, R., Goutierre, M., & Zugaro, M. (2016). Hippocampo-cortical coupling mediates memory consolidation during sleep. *Nature Neuroscience*, *19*, 959–964.

Mangiafico, S. (2020). rcompanion: Functions to Support Extension Education Program Evaluation. Retrieved from https://cran.r-project.org/package=rcompanion

Maquet, P., Laureys, S., Peigneux, P., Fuchs, S., Petiau, C., Phillips, C., … Cleeremans, A. (2000). Experience-dependent changes in changes in cerebral activation during human REM sleep. *Nature Neuroscience*, *3*(8), 831–836.

Maquet, P., Schwartz, S., Passingham, R., & Frith, C. (2003). Sleep-related consolidation of a visuomotor skill: Brain mechanisms as assessed by functional magnetic resonance imaging. *Journal of Neuroscience*, *23*(4), 1432–1440.

Marr, D., Willshaw, D., & McNaughton, B. (1971). Simple Memory: A Theory for Archicortex. *Philosophical Transactions of the Royal Society of London*, *262*(841), 23–81.

Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods*, *9*(2), 147–163.

May, C. P. (1999). Synchrony effects in cognition: The costs and a benefit. *Psychonomic Bulletin & Review*, *6*, 142–147.

May, C. P., Hasher, L., & Foong, N. (2005). Implicit memory, age, and time of day: Paradoxical priming effects. *Psychological Science*, *16*(2), 96–100.

May, C. P., Hasher, L., & Stoltzfus, E. R. (1993). Optimal Time of Day and the Magnitude of Age Differences in Memory. *Psychological Science*, *4*(5), 326–330.

Mazzarello, P. (2000). What dreams may come? *Nature*, *408*, 523.

Mazzoni, G., Gori, S., Formicola, G., Gneri, C., Massetani, R., Murri, L., & Salzarulo, P. (1999). Word recall correlates with sleep cycles in elderly subjects. *Journal of Sleep Research*, *8*(3), 185–188.

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.

McClung, C. R. (2006). Plant circadian rhythms. *The Plant Cell*, *18*(4), 792–803.

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.

McKoon, G., Ratcliff, R., & Dell, G. S. (1986). A Critical Evaluation of the Semantic-Episodic Distinction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(2), 295–306.

Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review*, *69*, 220–232.

Mednick, S., Nakayama, K., & Stickgold, R. (2003). Sleep-dependent learning: A nap is as good as a night. *Nature Neuroscience*, *6*(7), 697–698.

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology*, *110*(1), 40–48.

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc*, *36*(4), 630–633.

Mölle, M., Bergmann, T. O., Marshall, L., & Born, J. (2011). Fast and Slow Spindles during the Sleep Slow Oscillation: Disparate Coalescence and Engagement in Memory Processing. *Sleep*, *34*(10), 1411–1421.

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.

Monaghan, P., Sio, U. N., Lau, S. W., Woo, H. K., Linkenauger, S. A., & Ormerod, T. C. (2015). Sleep promotes analogical transfer in problem solving. *Cognition*, *143*, 25–30.

Moser, D., Anderer, P., Gruber, G., Parapatics, S., Loretz, E., Boeck, M., … Dorffner, G. (2009). Sleep Classification According to AASM and Rechtschaffen & Kales: Effects on Sleep Scoring Parameters. *Sleep*, *32*(2), 139–149.

Murphy, M., Stickgold, R., Parr, M. E., Callahan, C., & Wamsley, E. J. (2018). Recurrence of task-related electroencephalographic activity during post-training quiet rest and sleep. *Scientific Reports*, *8*(1), 1–10.

Nádasdy, Z., Hirase, H., Czurkó, A., Csicsvari, J., & Buzsáki, G. (1999). Replay and time compression of recurring spike sequences in the hippocampus. *Journal of Neuroscience*, *19*(21), 9497–9507.

Nakashiba, T., Buhl, D. L., McHugh, T. J., & Tonegawa, S. (2009). Hippocampal CA3 Output Is Crucial for Ripple-Associated Reactivation and Consolidation of Memory. *Neuron*, *62*(6), 781–787.

Navarrete, M., Schneider, J., Ngo, H.-V. V, Valderrama, M., Casson, A. J., & Lewis, P. A. (2020).

Examining the optimal timing for closed-loop auditory stimulation of slow-wave sleep in young and older adults. *Sleep*, *43*(6), zsz315.

Navarrete, M., Valderrama, M., & Lewis, P. A. (2020). The role of slow-wave sleep rhythms in the cortical-hippocampal loop for memory consolidation. *Current Opinion in Behavioral Sciences*, *32*, 102–110.

Neckelmann, D., & Ursin, R. (1993). Sleep Stages and EEG Power Spectrum in Relation to Acoustical Stimulus Arousal Threshold in the Rat. *Sleep*, *16*(5), 467–477.

Nemeth, D., Janacsek, K., Londe, Z., Ullman, M. T., Howard, D. V., & Howard, J. H. (2010). Sleep has no critical role in implicit motor sequence learning in young and old adults. *Experimental Brain Research*, *201*, 351–358.

Nere, A., Hashmi, A., Cirelli, C., & Tononi, G. (2013). Sleep-Dependent Synaptic Down-Selection (I): Modeling the Benefits of Sleep on Memory Consolidation and Integration. *Frontiers in Neurology*, *4*, 143.

Newbury, C. R., & Monaghan, P. (2019). When does sleep affect veridical and false memory consolidation? A meta-analysis. *Psychonomic Bulletin & Review*, *26*(2), 387–400.

Niiyama, Y., Fujiwara, R., Satoh, N., & Hishikawa, Y. (1994). Endogenous components of event-related potential appearing during NREM stage 1 and REM sleep in man. *International Journal of Psychophysiology*, *17*(2), 165–174.

Nordby, H., Hugdahl, K., Stickgold, R., Bronnick, K. S., & Hobson, J. A. (1996). Event-related potentials (ERPs) to deviant auditory stimuli during sleep and waking. *NeuroReport*, *7*(5), 1082–1086.

O'Keefe, J., Nadel, L., & Willner, J. (1979). Tuning out irrelevancy? Comments on Solomon's temporal mapping view of the hippocampus. *Psychological Bulletin*, *86*(6), 1280–1289.

O'Neill, J., Senior, T. J., Allen, K., Huxter, J. R., & Csicsvari, J. (2008). Reactivation of experience-dependent cell assembly patterns in the hippocampus. *Nature Neuroscience*, *11*, 209–215.

O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review*, *108*(2), 311–345.

Olejnik, S., & Algina, J. (2003). Generalized Eta and Omega Squared Statistics: Measures of Effect Size for Some Common Research Designs. *Psychological Methods*, *8*(4), 434–447.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, Article ID: 156869.

Oudiette, D., & Paller, K. A. (2013). Upgrading the sleeping brain with targeted memory reactivation. *Trends in Cognitive Sciences*, *17*(3), 142–149.

Pardilla-Delgado, E., & Payne, J. D. (2017). The impact of sleep on true and false memory across long delays. *Neurobiology of Learning and Memory*, *137*, 123–133.

Patanaik, A., Ong, J. L., Gooley, J. J., Ancoli-Israel, S., & Chee, M. W. L. (2018). An end-to-end framework for real-time automatic sleep stage classification. *Sleep*, *41*(5), zsy041.

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. *Journal of Neuroscience*, *9*(8), 2907–2918.

Payne, J. D., Chambers, A., & Kensinger, E. (2012). Sleep promotes lasting changes in selective memory for emotional scenes. *Frontiers in Integrative Neuroscience*, *6*, 108.

Payne, J. D., Schacter, D. L., Propper, R. E., Huang, L. W., Wamsley, E. J., Tucker, M. A., … Stickgold, R. (2009). The role of sleep in false memory formation. *Neurobiology of Learning and Memory*, *92*(3), 327–334.

Peigneux, P., Laureys, S., Delbeuck, X., & Maquet, P. (2001). Sleeping brain, learning brain. The role of sleep for memory systems. *NeuroReport*, *12*(18), A111–A124.

Peigneux, P., Laureys, S., Fuchs, S., Collette, F., Perrin, F., Reggers, J., … Maquet, P. (2004). Are spatial memories strengthened in the human hippocampus during slow wave sleep? *Neuron*, *44*(3), 535–545.

Peigneux, P., Laureys, S., Fuchs, S., Destrebecqz, A., Collette, F., Delbeuck, X., … Maquet, P. (2003). Learned material content and acquisition level modulate cerebral reactivation during posttraining rapid-eye-movements sleep. *NeuroImage*, *20*(1), 125–134.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203.

Pennartz, C. M. A., Lee, E., Verheul, J., Lipa, P., Barnes, C. A., & McNaughton, B. L. (2004). The ventral striatum in off-line processing: Ensemble reactivation during sleep and modulation by hippocampal ripples. *Journal of Neuroscience*, *24*(29), 6446–6456.

Pereira, S. I. R., Beijamini, F., Weber, F. D., Vincenzi, R. A., da Silva, F. A. C., & Louzada, F. M. (2017). Tactile stimulation during sleep alters slow oscillation and spindle densities but not motor skill. *Physiology and Behavior*, *169*, 59–68.

Petzka, M., Charest, I., Balanos, G. M., & Staresina, B. P. (2020). Does sleep-dependent consolidation favour weak memories? *PsyArXiv*. https://doi.org/10.31234/osf.io/q4wnv

Peyrache, A., Khamassi, M., Benchenane, K., Wiener, S. I., & Battaglia, F. P. (2009). Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nature Neuroscience*, *12*, 919–926.

Peyrache, A., & Seibt, J. (2020). A mechanism for learning with sleep spindles. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1799), 20190230.

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.

Plucker, J. A. (1999). Is the proof in the pudding? Reanalyses of Torrance's (1958 to present) longitudinal data. *Creativity Research Journal*, *12*(2), 103–114.

Poe, G. R., Nitz, D. A., McNaughton, B. L., & Barnes, C. A. (2000). Experience-dependent phase-reversal of hippocampal neuron firing during REM sleep. *Brain Research*, *855*(1), 176–180.

Poe, G. R., Walsh, C. M., & Bjorness, T. E. (2010). Cognitive neuroscience of sleep. *Progress in Brain Research*, *185*, 1–19.

Posner, M. I. (1973). *Cognition: An Introduction*. Glenview, IL: Scott, Foresman.

Prabhakaran, R., Green, A. E., & Gray, J. R. (2014). Thin slices of creativity: Using single-word utterances to assess creative cognition. *Behavior Research Methods*, *46*(3), 641–659.

Pratt, H., Berlad, I., & Lavie, P. (1999). "Oddball" event-related potentials and information processing during REM and non-REM sleep. *Clinical Neurophysiology*, *110*(1), 53–61.

Preston, A. R., Shrager, Y., Dudukovic, N. M., & Gabrieli, J. D. E. (2004). Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus*, *14*(2), 148–152.

Purcell, S. M., Manoach, D. S., Demanuele, C., Cade, B. E., Mariani, S., Cox, R., … Stickgold, R. (2017). Characterizing sleep spindles in 11,630 individuals from the National Sleep Research Resource. *Nature Communications*, *8*, 15930.

Qin, Y.-L., Mcnaughton, B. L., Skaggs, W. E., & Barnes, C. A. (1997). Memory reprocessing in corticocortical and hippocampocortical neuronal ensembles. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *352*(1360), 1525–1533.

R Core Team. (2020). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Review*, *93*, 681–766.

Rasch, B., Büchel, C., Gais, S., & Born, J. (2007). Odor cues during Slow-Wave Sleep prompt

declarative memory consolidation. *Science*, *315*, 1426–1429.

Rauchs, G., Bertran, F., Guillery-Girard, B., Desgranges, B., Kerrouche, N., Denise, P., … Eustache, F. (2004). Consolidation of strictly episodic memories mainly requires rapid eye movement sleep. *Sleep*, *27*(3), 395–401.

Renoult, L., Irish, M., Moscovitch, M., & Rugg, M. D. (2019). From Knowing to Remembering: The Semantic–Episodic Distinction. *Trends in Cognitive Sciences*, *23*(12), 1041–1057.

Ridding, M. C., & Flavel, S. C. (2006). Induction of plasticity in the dominant and non-dominant motor cortices of humans. *Experimental Brain Research*, *171*(4), 551–557.

Rihm, J. S., & Rasch, B. (2015). Replay of conditioned stimuli during late REM and stage N2 sleep influences affective tone rather than emotional memory strength. *Neurobiology of Learning and Memory*, *122*, 142–151.

Ritter, S. M., Strick, M., Bos, M. W., Van Baaren, R. B., & Dijksterhuis, A. (2012). Good morning creativity: Task reactivation during sleep enhances beneficial effect of sleep on creative performance. *Journal of Sleep Research*, *21*(6), 643–647.

Robertson, E. M., Pascual-Leone, A., & Press, D. Z. (2004). Awareness Modifies the Skill-Learning Benefits of Sleep. *Current Biology*, *14*(3), 208–212.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814.

Ruch, S., Markes, O., Duss, S. B., Oppliger, D., Reber, T. P., Koenig, T., … Henke, K. (2012). Sleep stage II contributes to the consolidation of declarative memories. *Neuropsychologia*, *50*(10), 2389–2396.

Rudoy, J. D., Voss, J. L., Westerberg, C. E., & Paller, K. A. (2009). Strengthening Individual Memories by Reactivating Them During Sleep. *Science*, *326*(5956), 1079.

Runco, M. A., & Jaeger, G. J. (2012). The Standard Definition of Creativity. *Creativity Research Journal*, *24*(1), 92–96.

Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. *Behavioral Ecology*.

Saletin, J. M., Goldstein, A. N., & Walker, M. P. (2011). The Role of Sleep in Directed Forgetting and Remembering of Human Memories. *Cerebral Cortex*, *21*(11), 2534–2541.

Sallinen, M., Kaartinen, J., & Lyytinen, H. (1996). Processing of auditory stimuli during tonic and phasic periods of REM sleep as revealed by event-related brain potentials. *Journal of Sleep*

Research, *5*(4), 220–228.

Sanders, K. E. G., Osburn, S., Paller, K. A., & Beeman, M. (2019). Targeted Memory Reactivation During Sleep Improves Next-Day Problem Solving. *Psychological Science*, *30*(11), 1616–1624.

Sandoval, M., Leclerc, J. A., & Gómez, R. L. (2017). Words to Sleep On: Naps Facilitate Verb Generalization in Habitually and Nonhabitually Napping Preschoolers. *Child Development*, *88*(5), 1615–1628.

Sara, S. J. (2017). Sleep to Remember. *Journal of Neuroscience*, *37*(3), 457–463.

Schabus, M., Gruber, G., Parapatics, S., Sauter, C., Klösch, G., Anderer, P., … Zeitlhofer, J. (2004). Sleep Spindles and Their Significance for Declarative Memory Consolidation. *Sleep*, *27*(8), 1479–1485.

Schapiro, A. C., McDevitt, E. A., Rogers, T. T., Mednick, S. C., & Norman, K. A. (2018). Human hippocampal replay during rest prioritizes weakly learned information and predicts memory performance. *Nature Communications*, *9*, 3920.

Schechtman, E., Witkowski, S., Lampe, A., Wilson, B. J., & Paller, K. A. (2020). Targeted memory reactivation during sleep boosts intentional forgetting of spatial locations. *Scientific Reports*, *10*, 2327.

Schlichting, M. L., Guarino, K. F., Schapiro, A. C., Turk-Browne, N. B., & Preston, A. R. (2017). Hippocampal structure predicts statistical learning and associative inference abilities during development. *Journal of Cognitive Neuroscience*, *29*(1), 37–51.

Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature Communications*, *6*, 8151.

Schlögl, A., Keinrath, C., Zimmermann, D., Scherer, R., Leeb, R., & Pfurtscheller, G. (2007). A fully automated correction method of EOG artifacts in EEG recordings. *Clinical Neurophysiology*, *118*(1), 98–104.

Schönauer, M., Alizadeh, S., Jamalabadi, H., Abraham, A., Pawlizki, A., & Gais, S. (2017). Decoding material-specific memory reprocessing during sleep in humans. *Nature Communications*, *8*, 15404.

Schönauer, M., Brodt, S., Pöhlchen, D., Breßmer, A., Danek, A. H., & Gais, S. (2018). Sleep Does Not Promote Solving Classical Insight Problems and Magic Tricks. *Frontiers in Human Neuroscience*, *12*, 72.

Schönauer, M., Geisler, T., & Gais, S. (2014). Strengthening Procedural Memories by Reactivation in Sleep. *Journal of Cognitive Neuroscience*, *26*(1), 143–153.

Schönbrodt, F. D., Wagenmakers, E. J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*(2), 322–339.

Schreiner, T., Doeller, C. F., Jensen, O., Rasch, B., & Staudigl, T. (2018). Theta Phase-Coordinated Memory Reactivation Reoccurs in a Slow-Oscillatory Rhythm during NREM Sleep. *Cell Reports*, *25*(2), 296–301.

Schreiner, T., Lehmann, M., & Rasch, B. (2015). Auditory feedback blocks memory benefits of cueing during sleep. *Nature Communications*, *6*, 8729.

Schreiner, T., & Rasch, B. (2015). Boosting Vocabulary Learning by Verbal Cueing During Sleep. *Cerebral Cortex*, *25*(11), 4169–4179.

Schreiner, T., & Staudigl, T. (2020). Electrophysiological signatures of memory reactivation in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*, 20190293.

Scullin, M. K., & Gao, C. (2018). Dynamic Contributions of Slow Wave Sleep and REM Sleep to Cognitive Longevity. *Current Sleep Medicine Reports*, *4*, 284–293.

Shanahan, L. K., Gjorgieva, E., Paller, K. A., Kahnt, T., & Gottfried, J. A. (2018). Odor-evoked category reactivation in human ventromedial prefrontal cortex during sleep promotes memory consolidation. *ELife*, *7*, e39681.

Shen, J., Kudrimoti, H., McNaughton, B. L., & Barnes, C. (1998). Reactivation of neuronal ensembles in hippocampal dentate gyrus during sleep after spatial experience. *Journal of Sleep Research*, *7*(S1), 6–16.

Shimizu, R. E., Connolly, P. M., Cellini, N., Armstrong, D. M., Hernandez, L. T., Estrada, R., … Simons, S. B. (2018). Closed-Loop Targeted Memory Reactivation during Sleep Improves Spatial Navigation. *Frontiers in Human Neuroscience*, *12*, 28.

Siegel, J. M. (2008). Do all animals sleep? *Trends in Neurosciences*, *31*(4), 208–213.

Silber, M. H., Ancoli-Israel, S., Bonnet, M. H., Chokroverty, S., Grigg-Damberger, M. M., Hirshkowitz, M., … Iber, C. (2007). The Visual Scoring of Sleep in Adults. *Journal of Clinical Sleep Medicine*, *3*(2).

Simon, K. C. N. S., Gómez, R. L., & Nadel, L. (2018). Losing memories during sleep after targeted memory reactivation. *Neurobiology of Learning and Memory*, *151*, 10–17.

Simon, K. C. N. S., Werchan, D., Goldstein, M. R., Sweeney, L., Bootzin, R. R., Nadel, L., & Gómez, R. L. (2017). Sleep confers a benefit for retention of statistical language learning in 6.5 month old infants. *Brain and Language*, *167*, 3–12.

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2020). afex: Analysis of Factorial Experiments. Retrieved from https://cran.r-project.org/package=afex

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*(2), 159–166.

Sio, U. N., & Ormerod, T. C. (2009). Does Incubation Enhance Problem Solving? A Meta-Analytic Review. *Psychological Bulletin*, *135*(1), 94–120.

Skaggs, W. E., & McNaughton, B. L. (1996). Replay of Neuronal Firing Sequences in Rat Hippocampus During Sleep Following Spatial Experience. *Science*, *271*(5257), 1870–1873.

Skaggs, W. E., McNaughton, B. L., Wilson, M. A., & Barnes, C. A. (1996). Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus*, *6*(2), 149–172.

Smith, C. (1993). REM Sleep and Learning: Some Recent Findings. In A. Moffitt, M. Kramer, & R. Hoffmann (Eds.), *The Functions of Dreaming* (pp. 341–361). Albany, NY: State University of New York Press.

Smith, C. (1995). Sleep states and memory processes. *Behavioural Brain Research*, *69*(1–2), 137–145.

Smith, C. (2001). Sleep states and memory processes in humans: Procedural versus declarative memory systems. *Sleep Medicine Reviews*, *5*(6), 491–506.

Smith, C. (2011). Sleep states and memory processing in rodents: A review. *Sleep Medicine Clinics*, *6*(1), 59–70.

Smith, C., & Smith, D. (2003). Ingestion of Ethanol Just Prior to Sleep Onset Impairs Memory for Procedural but not Declarative Tasks. *Sleep*, *26*(2), 185–191.

Smith, C., & Weeden, K. (1990). Post training REMs coincident auditory stimulation enhances memory in humans. *Psychiatric Journal of the University of Ottawa*, *15*(2), 85–90.

Song, S., Howard, J. H., & Howard, D. V. (2007). Sleep does not benefit probabilistic motor sequence learning. *Journal of Neuroscience*, *27*(46), 12475–12483.

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

Spalding, K. N., Schlichting, M. L., Zeithamova, D., Preston, A. R., Tranel, D., Duff, M. C., & Warren, D. E. (2018). Ventromedial prefrontal cortex is necessary for normal associative inference and memory integration. *Journal of Neuroscience*, *38*(15), 3767–3775.

Spencer, R. M. C., Sunm, M., & Ivry, R. B. (2006). Sleep-Dependent Consolidation of Contextual Learning. *Current Biology*, *16*(10), 1001–1005.

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, *82*(3), 171–177.

Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences*, *93*(24), 13515–13522.

Stein, M. I. (1953). Creativity and Culture. *The Journal of Psychology*, *36*, 311–322.

Sternberg, R. J., & Lubart, T. I. (1996). Investing in creativity. *American Psychologist*, *51*(7), 677–688.

Sterpenich, V., Schmidt, C., Albouy, G., Matarazzo, L., Vanhaudenhuyse, A., Boveroux, P., … Maquet, P. (2014). Memory Reactivation during Rapid Eye Movement Sleep Promotes Its Generalization and Integration in Cortical Stores. *Sleep*, *37*(6), 1061–1075.

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, *437*(7063), 1272–1278.

Stickgold, R., Scott, L., Rittenhouse, C., & Hobson, J. A. (1999). Sleep-Induced Changes in Associative Memory. *Journal of Cognitive Neuroscience*, *11*(2), 182–193.

Stickgold, R., & Walker, M. P. (2013). Sleep-dependent memory triage: Evolving generalization through selective processing. *Nature Neuroscience*, *16*, 139–145.

Stickgold, R., Whidbee, D., Schirmer, B., Patel, V., & Hobson, J. A. (2000). Visual discrimination task improvement: A multi-step process occurring during sleep. *Journal of Cognitive Neuroscience*, *12*(2), 246–254.

Storm, B. C., & Angello, G. (2010). Overcoming Fixation: Creative Problem Solving and Retrieval-Induced Forgetting. *Psychological Science*, *21*(9), 1263–1265.

Storm, B. C., Ditta, A. S., & George, T. (2020). Memory. In S. Pritzker & M. A. Runco (Eds.), *Encyclopedia of Creativity* (3rd ed., pp. 116–120). Academic Press.

Studte, S., Bridger, E., & Mecklinger, A. (2017). Sleep spindles during a nap correlate with post sleep memory performance for highly rewarded word-pairs. *Brain and Language*, *167*, 28–35.

Sweegers, C. C. G., & Talamini, L. M. (2014). Generalization from episodic memories across time: A route for semantic knowledge acquisition. *Cortex*, *59*, 49–61.

Takahara, M., Nittono, H., & Hori, T. (2006). Effect of Voluntary Attention on Auditory Processing During REM Sleep. *Sleep*, *29*(7), 975–982.

Tambini, A., Berners-Lee, A., & Davachi, L. (2017). Brief targeted memory reactivation during the awake state enhances memory stability and benefits the weakest memories. *Scientific Reports*, *7*, 15325.

Tambini, A., & Davachi, L. (2019). Awake Reactivation of Prior Experiences Consolidates Memories and Biases Cognition. *Trends in Cognitive Sciences*, *23*(10), 876–890.

Teasdale, J. D., & Fogarty, S. J. (1979). Differential effects of induced mood on retrieval of pleasant and unpleasant events from episodic memory. *Journal of Abnormal Psychology*, *88*(3), 248–257.

Tobler, I., Franken, P., Trachsel, L., & Borbély, A. A. (1992). Models of sleep regulation in mammals. *Journal of Sleep Research*, *1*(2), 125–127.

Tononi, G., & Cirelli, C. (2003). Sleep and synaptic homeostasis: A hypothesis. *Brain Research Bulletin*, *62*(2), 143–150.

Tononi, G., & Cirelli, C. (2014). Sleep and the Price of Plasticity: From Synaptic and Cellular Homeostasis to Memory Consolidation and Integration. *Neuron*, *81*(1), 12–34.

Tononi, G., & Cirelli, C. (2016). Sleep and Synaptic Down-Selection. In G. Buzsáki & Y. Christen (Eds.), *Micro-, Meso- and Macro-Dynamics of the Brain. Research and Perspectives in Neurosciences* (pp. 99–106). Springer.

Torchiano, M. (2020). effsize: Efficient Effect Size Computation. Retrieved from https://cran.r-project.org/package=effsize

Torrance, E. P. (1981). Predicting the Creativity of Elementary School Children (1958-80) —and the Teacher Who "Made a Difference." *Gifted Child Quarterly*, *25*(2), 55–62.

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., … Morris, R. G. M. (2007). Schemas and memory consolidation. *Science*, *316*(5821), 76–82.

Tulving, E. (1983). *Elements of Episodic Memory*. London: Oxford University Press.

Ulrich, D. (2016). Sleep Spindles as Facilitators of Memory Formation and Learning. *Neural Plasticity*, *2016*, 1796715.

Van Der Werf, Y. D., Van Der Helm, E., Schoonheim, M. M., Ridderikhoff, A., & Van Someren, E. J. W. (2009). Learning by observation requires an early sleep window. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(45), 18926–18930.

van Dongen, H. P. A., Vitellaro, K. M., & Dinges, D. F. (2005). Individual Differences in Adult Human Sleep and Wakefulness: Leitmotif for a Research Agenda. *Sleep*, *28*(4), 479–498.

Van Ormer, E. B. (1933). Sleep and retention. *Psychological Bulletin*, *30*(6), 415–439.

Vartanian, O., Bouak, F., Caldwell, J. L., Cheung, B., Cupchik, G., Jobidon, M.-E., … Smith, I. (2014). The effects of a single night of sleep deprivation on fluency and prefrontal cortex function during divergent thinking. *Frontiers in Human Neuroscience*, *8*, 214.

Vartanian, O., Jobidon, M.-E., Bouak, F., Nakashima, A., Smith, I., Lam, Q., & Cheung, B. (2013). Working memory training is associated with lower prefrontal cortex activation in a divergent thinking task. *Neuroscience*, *236*, 186–194.

Veale, J. F. (2014). Edinburgh Handedness Inventory - Short Form: A revised version based on confirmatory factor analysis. *Laterality: Asymmetries of Body, Brain and Cognition*, *19*(2), 164–177.

Veasey, S. C., Valladares, O., Fenik, P., Kapfhamer, D., Sanford, L., Benington, J., & Bucan, M. (2000). An automated system for recording and analysis of sleep in mice. *Sleep*, *23*(8), 1025–1040.

Verleger, R., Rose, M., Wagner, U., Yordanova, J., & Kolev, V. (2013). Insights into sleep's role for insight: Studies with the number reduction task. *Advances in Cognitive Psychology*, *9*(4), 160–172.

Vigário, R. N. (1997). Extraction of ocular artefacts from EEG using independent component analysis. *Electroencephalography and Clinical Neurophysiology*, *103*(3), 395–404.

Volkmann, J., Schnitzler, A., Witte, O. W., & Freund, H.-J. (1998). Handedness and Asymmetry of Hand Representation in Human Motor Cortex. *Journal of Neurophysiology*, *79*(4), 2149–2154.

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., … Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*, 35–57.

Wagner, U., Gais, S., & Born, J. (2001). Emotional memory formation is enhanced across sleep intervals with high amounts of rapid eye movement sleep. *Learning and Memory*, *8*(2), 112–119.

Wagner, U., Gais, S., Haider, H., Verleger, R., & Born, J. (2004). Sleep inspires insight. *Nature*, *427*(6972), 352–355.

Walker, M. P., Brakefield, T., Hobson, J. A., & Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature*, *425*(6958), 616–620.

Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A., & Stickgold, R. (2002). Practice with sleep makes perfect: Sleep-dependent motor skill learning. *Neuron*, *35*(1), 205–211.

Walker, M. P., Brakefield, T., Seidman, J., Morgan, A., Hobson, J. A., & Stickgold, R. (2003). Sleep

and the time course of motor skill learning. *Learning and Memory*, *10*(4), 275–284.

Walker, M. P., Liston, C., Hobson, J. A., & Stickgold, R. (2002). Cognitive flexibility across the sleep–wake cycle: REM-sleep enhancement of anagram problem solving. *Cognitive Brain Research*, *14*, 317–324.

Walker, M. P., Stickgold, R., Alsop, D., Gaab, N., & Schlaug, G. (2005). Sleep-dependent motor memory plasticity in the human brain. *Neuroscience*, *133*(4), 911–917.

Wang, B., Antony, J. W., Lurie, S., Brooks, P. P., Paller, K. A., & Norman, K. A. (2019). Targeted Memory Reactivation during Sleep Elicits Neural Signals Related to Learning Content. *Journal of Neuroscience*, *39*(34), 6728–6736.

Werchan, D. M., & Gómez, R. L. (2013). Generalizing memories over time: Sleep and reinforcement facilitate transitive inference. *Neurobiology of Learning and Memory*, *100*, 70–76.

Werchan, D. M., & Gómez, R. L. (2014). Wakefulness (Not Sleep) Promotes Generalization of Word Learning in 2.5-Year-Old Children. *Child Development*, *85*(2), 429–436.

Werth, E., Achermann, P., Dijk, D. J., & Borbély, A. A. (1997). Spindle frequency activity in the sleep EEG: Individual differences and topographic distribution. *Electroencephalography and Clinical Neurophysiology*, *103*(5), 535–542.

Wieth, M. B., & Zacks, R. T. (2011). Time of day effects on problem solving: When the non-optimal is optimal. *Thinking & Reasoning*, *17*(4), 387–401.

Wilhelm, I., Diekelmann, S., Molzow, I., Ayoub, A., Mölle, M., & Born, J. (2011). Sleep Selectively Enhances Memory Expected to Be of Future Relevance. *Journal of Neuroscience*, *31*(5), 1563–1569.

Wilhelm, I., Rose, M., Imhof, K. I., Rasch, B., Büchel, C., & Born, J. (2013). The sleeping child outplays the adult's capacity to convert implicit into explicit knowledge. *Nature Neuroscience*, *16*, 391–393.

Williams, S. E., Cumming, J., Ntoumanis, N., Nordin-Bates, S. M., Ramsey, R., & Hall, C. (2012). Further validation and development of the Movement Imagery Questionnaire. *Journal of Sport & Exercise Psychology*, *34*, 621–646.

Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of Hippocampal Ensemble Memories During Sleep. *Science*, *265*(5172), 676–679.

Wimmer, F., Hoffmann, R. F., Bonato, R. A., & Moffitt, A. R. (1992). The effects of sleep deprivation on divergent thinking and attention processes. *Journal of Sleep Research*, *1*(4), 223–230.

Witt, K., Margraf, N., Bieber, C., Born, J., & Deuschl, G. (2010). Sleep consolidates the effector-

independent representation of a motor skill. *Neuroscience*, *171*(1), 227–234.

Wixted, J. T. (2004). The Psychology and Neuroscience of Forgetting. *Annual Review of Psychology*, *55*(1), 235–269.

Yang, L., Hasher, L., & Wilson, D. E. (2007). Synchrony effects in automatic and controlled retrieval. *Psychonomic Bulletin & Review*, *14*, 51–56.

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.

Yasenkov, R., & Deboer, T. (2010). Circadian Regulation of Sleep and the Sleep EEG under Constant Sleep Pressure in the Rat. *Sleep*, *33*(5), 631–641.

Yordanova, J., Kolev, V., Verleger, R., Bataghva, Z., Born, J., & Wagner, U. (2008). Shifting from implicit to explicit knowledge: Different roles of early- and late-night sleep. *Learning and Memory*, *15*, 508–515.

Yotsumoto, Y., Sasaki, Y., Chan, P., Vasios, C. E., Bonmassar, G., Ito, N., … Watanabe, T. (2009). Location-Specific Cortical Activation Changes during Sleep after Training for Perceptual Learning. *Current Biology*, *19*(15), 1278–1282.

Zeithamova, D., Dominick, A. L., & Preston, A. R. (2012). Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron*, *75*(1), 168–179.

Zeithamova, D., & Preston, A. R. (2010). Flexible memories: Differential roles for medial temporal lobe and prefrontal cortex in cross-episode binding. *Journal of Neuroscience*, *30*(44), 14676–14684.

Zhang, H., Fell, J., & Axmacher, N. (2018). Electrophysiological mechanisms of human memory consolidation. *Nature Communications*, *9*, 4103.

Zhang, J., Yetton, B., Whitehurst, L. N., Naji, M., & Mednick, S. C. (2020). The Effect of Zolpidem on Memory Consolidation Over a Night of Sleep. *Sleep*, zsaa084.

Züst, M. A., Ruch, S., Wiest, R., & Henke, K. (2019). Implicit Vocabulary Learning during Sleep Is Bound to Slow-Wave Peaks. *Current Biology*, *29*(4), 541-553.e7.

Appendices

Appendix A: Verb Generation Task instructions

Instructions Part 1: Common

You are about to begin Part 1. When you see the noun, generate a verb according to the following instructions:

Give a very common or typical verb response to the noun. By “common,” we mean a verb that is clearly related to the noun, and very often used in association with the noun. A verb that would probably come to almost everyone’s mind when they read the noun.

Instructions Part 2: Random

You are about to begin Part 2. When you see the noun, generate a verb according to the following instructions:

Give a very unusual or random verb response to the noun. By “unusual,” we mean a verb that is unrelated to the noun, and not used in association with the noun. Remember: you should only type in ONE verb. On the next page, you will see the noun. Think of the verb response you want to give, and once you have decided, type it in and press ENTER.

Instructions Part 3: Creative

You are about to begin Part 3. When you see the noun, generate a verb according to the following instructions:

Give a very creative or original verb response to the noun. By “creative,” we mean a verb that is clearly related to the noun, but also rarely used in association with the noun. A verb that is unusual, but still related to the noun.

Appendix B: Verb Generation Task noun cue list

Table A1. *List of nouns used in the Verb Generation Task, which is a subset of the nouns used in Prabhakaran et al.* (2014)*. Constraint level, most common response, and frequency of most common response were determined by Prabhakaran and colleagues in an independent sample.*

Item number	Noun	Constraint level	Most common response	Frequency of most common response
1	leaf	low	fall	.437
2	taxi	low	drive	.380
3	house	low	live	.380
4	rock	high	throw	.521
5	card	high	play	.465
6	oath	high	take	.521
7	street	low	walk	.408
8	lamp	low	light	.423
9	canoe	low	paddle	.408
10	home	low	live	.352
11	office	high	work	.606
12	belt	high	wear	.535
13	finger	high	point	.521
14	blade	high	cut	.521
15	oven	low	cook	.380
16	fist	low	punch	.408
17	cart	high	push	.493
18	tongue	low	lick	.380
19	shovel	high	dig	.521
20	dish	low	wash	.352
21	pill	high	take	.535
22	flower	low	smell	.437
23	note	high	write	.507
24	church	low	pray	.338
25	boot	low	wear	.423
26	snow	low	shovel	.380
27	feet	high	walk	.521
28	paper	low	write	.408
29	artist	high	paint	.479
30	cannon	high	shoot	.493
31	tune	low	play	.338
32	poem	high	read	.465
33	drum	low	play	.423
34	bucket	low	fill	.366
35	cafe	low	eat	.380
36	store	low	buy	.437
37	muscle	low	flex	.380
38	couch	high	sit	.563
39	drug	high	take	.592

40	rose	high	smell	.465
41	soup	high	eat	.507
42	letter	low	write	.423
43	pillow	high	sleep	.592
44	tool	high	use	.479
45	debt	high	pay	.606
46	music	low	play	.437
47	ring	low	wear	.380
48	hole	high	dig	.592
49	horn	high	blow	.606
50	golf	high	play	.592
51	manual	high	read	.577
52	infant	low	cry	.352
53	clay	low	mold	.352
54	hair	low	comb	.352
55	baby	high	cry	.507
56	money	high	spend	.479
57	phone	low	call	.408
58	glass	low	break	.408
59	pan	low	cook	.380
60	grass	low	cut	.380
61	bread	high	eat	.479
62	sofa	high	sit	.535
63	ship	high	sail	.507