The influence of prior knowledge on memory consolidation

Alexander McGregor Reid

Doctor of Philosophy

University of York

Department of Psychology

September 2013

Abstract

Current research indicates that the process of memory consolidation, particularly during periods of sleep, can benefit memory in a variety of ways. Many of these behavioural changes are thought to arise from the integration of new information into pre-existing semantic memory. Despite this contention, it is not presently well understood how the relationship between new memories and prior knowledge will influence the behavioural outcomes of consolidation. To explore this issue the present thesis contains a number of experiments designed to measure the effects of consolidation on new memories that are either semantically compatible (congruent) or incompatible (incongruent) with prior knowledge. Measures of sleep physiology were also obtained in Experiments 2-5 to elucidate the potential mechanistic contributions of this state on memory consolidation. Using an adapted transitive inference paradigm Experiment 1 revealed evidence for the rapid assimilation of both congruent and incongruent information over 24 hours. However in Experiment 2, which featured congruent information only, these effects were not replicated, suggesting that consolidation may be more beneficial for the integration of more weakly encoded information. To explore this matter further Experiments 3-5 examined how semantic congruency may influence the recovery of new memories weakened by retroactive interference (RI) using a modified A-B, A-C list learning paradigm. Over the course of these investigations sleep was found to facilitate the recovery of semantically congruent memories from RI, thus negating the detrimental impact of both specific and non-specific forms of waking interference on memory. Incongruent information, however, was simultaneously forgotten suggesting a selective component to this consolidation-associated recovery. The sleep measures obtained across Experiments 2-5 were largely inconclusive although there was provisional evidence that sleep physiology may support the flexible expression of declarative memory (Experiment 2) and that rapid eye movement (REM) sleep may facilitate RI recovery (Experiment 3). These findings are interpreted in light of existing models of sleep-associated memory consolidation, integration and RI recovery.

Table of Contents

Contents

Abstract 2

Table of Contents 3

List of Tables 8

List of Figures 10

Acknowledgements 12

Dedication 13

Declaration 14

Chapter 1: Sleep, prior knowledge and dynamic memory consolidation 15

- 1.1 Introduction 15
- 1.2 Memory systems and consolidation 16
  - 1.2.1 Consolidation 17
  - 1.2.2 The standard model of memory consolidation 18
  - 1.2.3 Evidence for the standard model 19
  - 1.2.4 Multiple Trace Theory 22
  - 1.2.5 Conclusions 23
- 1.3 Sleep and memory consolidation 24
  - 1.3.1 Sleep characteristics 24
  - 1.3.2 Sleep dependent memory consolidation 26
  - 1.3.3 Models relating to sleep stage activity 26
  - 1.3.4 Models relating to field potential oscillations 29
  - 1.3.5 The issue of memory strength and consolidation 36
  - 1.3.6 Criticisms of an active role for sleep in memory consolidation 37
  - 1.3.7 Summary and conclusions 39
- 1.4 Sleep and memory transformation 40
  - 1.4.1 Memory assimilation 40
  - 1.4.2 Memory abstraction 42
  - 1.4.3 Abstraction and integration in parallel 44
- 1.5 Conclusions and thesis outline 46
  - 1.5.1 Research questions 46
  - 1.5.2 Experimental chapters 47

Chapter 2: Consolidation, prior knowledge and integration 49

- 2.1 Introduction 49
  - 2.1.1 Transitive inference and consolidation 49
  - 2.1.2 Congruency and consolidation 53

2.1.3 Experiment aims and predictions 54
2.2 Methods 57
2.2.1 Participants 57
2.2.2 Design 58
2.2.3 Stimuli 58
2.2.4 Procedure 61
2.3 Results 64
2.3.1 Group equivalency 64
2.3.2 Premises pair performance and the serial position effect 65
2.3.3 Transitive inferences and memory abstraction 67
2.3.3 Size generalisations and the semantic distance effect 71
2.4 Summary and Discussion 72
2.4.1 The issue of congruency and consolidation 73
2.4.2 Consolidation facilitates the suppression of the serial position effect 74
2.4.3 Consolidation facilitates the development of relational memory 76
2.4.4 Diverging routes of consolidation 78
2.4.5 Novel word generalisations 79
2.4.6 Conclusions 80

Chapter 3: Consolidation, sleep and memory integration 81
3.1 Introduction 81
3.1.1 Memory integration and sleep stages 81
3.1.2 Memory integration and sleep spindles 84
3.1.3 Experiment aims and predictions 85
3.2 Methods 86
3.2.1 Participants 86
3.2.2 Design 87
3.2.3 Stimuli 88
3.2.4 Procedure 88
3.2.5 PSG recordings and analysis 90
3.2.6 Other measures 92
3.3 Results 92
3.3.1 Group equivalency 93
3.3.2 Premise pair performance and the serial position effect 95
3.3.3 Transitive inferences and memory abstraction 96
3.3.4 Size generalisations and the semantic distance effect 98
3.3.5 Sleep stages and integration 100
3.3.6 Sleep spindles and integration 103
3.4 Summary and Discussion 106

3.4.1 The lack of integration in Experiment 2 107
3.4.2 Sleep mechanisms and premise memory 108
3.4.3 Sleep and the maintenance of transitive inference performance 109
3.4.4 Size generalisations 110
3.4.5 Possible circadian influences 111
3.4.6 Conclusions 112

Chapter 4: The role of prior knowledge and consolidation in the recovery and integration of memory. 113
4.1 Introduction 113
4.1.1 Consolidation and the recuperation of memory 114
4.1.2 Consolidation and the recovery of memory from interference 115
4.1.3 Experiment 3 aims and predictions 117
4.2 Experiment 3 Methods 122
4.2.1 Participants 122
4.2.2 Design 122
4.2.3 Stimuli 123
4.2.4 Procedure 125
4.2.5 Sleep recordings 129
4.3 Experiment 3 results 129
4.3.1 Group equivalency 130
4.3.2 Cued recall performance across A-B and A-C lists 132
4.3.3 2AFC size judgements and retroactive interference 135
4.3.4 2AFC generalisations and novel word size focus 137
4.3.5 Sleep measures 140
4.4 Experiment 3 summary and discussion 142
4.5 Experiment 4 146
4.6 Experiment 4 methods 147
4.6.1 Participants 147
4.6.2 Procedure 148
4.7 Experiment 4 results 148
4.7.1 Group equivalency 149
4.7.2 Cued recall performance across A-B and A-C lists 151
4.7.3 2AFC size judgements and retroactive interference 153
4.7.4 2AFC generalisations and novel word size focus 155
4.7.5 Sleep measures 158
4.8 Experiment 4 summary and discussion 159
4.9 Chapter 4 summary and discussion 161
4.9.1 RI recovery and prior knowledge 163
4.9.2 Theoretical accounts of RI 163

4.9.3 The potential role of sleep in mnemonic recovery and memory integration 165
4.9.5 Conclusions 167

Chapter 5: The role of sleep and wakefulness on mnemonic recovery and integration..... 168
5.1 Introduction 168
5.1.1 Potential consolidation-associated mechanisms of RI recovery and integration169
5.1.2 Alternative accounts 171
5.1.3 Experiment 5 aims and predictions 171
5.2 Methods 174
5.2.1 Participants 174
5.2.2 Design 175
5.2.3 Procedure 176
5.2.4 PSG recording and analysis 176
5.2.5 Other measures 177
5.3 Experiment 5 results 177
5.3.1 Group equivalency 177
5.3.2 Dealing with differences in alertness 180
5.3.3 Cued recall performance across A-B and A-C lists 181
5.3.4 2AFC size judgements and retroactive interference 184
5.3.5 2AFC generalisations and novel word size focus 186
5.3.6 Sleep stage correlations 189
5.3.7 Sleep spindle activity 190
5.4 Summary and Discussion 191
5.4.1 Spontaneous recovery and memory consolidation 192
5.4.2 The role of forgetting and wakefulness on memory 195
5.4.3 Conclusions 195

Chapter 6: Thesis summary and conclusions 197
6.1 Thesis summary 197
6.1.1 Thesis motivation 197
6.1.2 Chapter 2 198
6.1.3 Chapter 3 199
6.1.4 Chapter 4 200
6.1.5 Chapter 5 201
6.2 The influence of prior knowledge on memory consolidation 202
6.2.1 Overview 202
6.2.2 The impact of congruency 203
6.2.3 Specific contributions to the existing literature 206
6.4 The sleep associated mechanisms underlying consolidation 209

6.4.1 Overview 209
6.4.2 Potential contributions of sleep activity on memory consolidation........ 210
6.5 Weaknesses of the present thesis 211
6.5.1 Circadian confounds 211
6.5.2 The issue of multiple comparisons 211
6.5.3 Other limitations........ 211

6.6 Conclusions and future work 213
6.6.1 Main contributions of this thesis 213
6.6.2 Future work........ 215

Appendices........ 217
Appendix 1 217
Appendix 2 218
Appendix 3 219
Appendix 4 220
Appendix 5 221
Appendix 6 222
Appendix 7 223
Appendix 8 224
Appendix 9 225
Appendix 10 233
Appendix 11 234
Appendix 12 242
Appendix 13 243
Appendix 14 244
Appendix 15 245
Appendix 16 246
Appendix 17 247
Appendix 18 248
Appendix 19 249
Appendix 20 250
Appendix 21 251

References 252

List of Tables

Table 2. Experiment 1: Accuracy levels for central and end premises. 66
Table 3. Experiment 1: RT performance for central and end premises. 67
Table 4. Experiment 1: Training conformity levels (%) for transitive inferences. 68
Table 5. Experiment 1: RT performance (msec) for transitive inferences. 69
Table 6. Experiment 1: RTs and the size proximity of generalisations for congruent and incongruent novel words. 71
Table 7. Experiment 2: Trained items and their size equivalents used in the generalisation phase. 90
Table 8. Experiment 2: Training performance. 93
Table 9. Experiment 2: Self-reported sleep patterns. 94
Table 10. Experiment 2: Stanford Sleepiness measures. 94
Table 11. Experiment 2: Size generalisation measures. 99
Table 12. Experiment 2: sleep parameters. 100
Table 13. Experiment 2: Correlations between integration measures and time spent in different stages of sleep. 102
Table 14. Experiment 2: Sleep spindle measures across sleep stage and electrode site. ... 104
Table 15. Experiment 2: Spindle activity in relation to integration measures. 105
Table 16. Stimuli sets and size ratings used for A-B, A-C list learning. 124
Table 17. Experiment 3: real word animal items and ratings to be used in the generalisation phase. 125
Table 18. Experiment 3: Exit block performance. 131
Table 19. Experiment 3: Stanford Sleepiness Scale (SSS) measures. 132
Table 20. Experiment 3: Recall rates for correct and incorrect list responses and untrained intrusion. 134
Table 21. Experiment 3: 2AFC reaction time measures across Congruency and List type and Delay type. 137
Table 22. Experiment 3: Generalisation performance for across 24 hours. 138
Table 23. Experiment 3: Sleep parameters for participants in the Delayed condition (n=11). 141
Table 24. Experiment 4: Exit block performance at the end of training for both Congruency and List types. 149
Table 25. Experiment 4: Self-rated sleep measures. 150
Table 26. Experiment 4: Stanford Sleepiness Scale (SSS) measures at training and testing for the Immediate and Delayed conditions. 150
Table 27. Experiment 4: Accuracy (%) and Accurate RTs (msec) on the rapid response control task. 151
Table 28. Experiment 4: Recall rates for correct and incorrect list responses and untrained intrusion rates. 152
Table 29. Experiment 4: Reaction time differences between training and testing across both congruency and list conditions. 154
Table 30. Experiment 4: Generalisation performance across 24 hours. 156
Table 31. Experiment 4: Sleep parameters. 158
Table 32. Experiment 5: Congruent and Incongruent accuracy levels for list learning on the final block of training. 178
Table 33. Experiment 5: Stanford Sleepiness Scale (SSS) measures. 179
Table 34. Experiment 5: Accuracy (%) and Accurate RTs (msec) on the rapid response control task. 180
Table 35. Experiment 5: Recall rates for correct and incorrect list responses and untrained intrusions rates. 182
Table 36. Experiment 5: 2AFC reaction time measures. 186
Table 37. Experiment 5: Generalisation performance in the wake and sleep conditions... 187
Table 38. Experiment 5: Sleep parameters. 189
Table 39. Experiment 5: Spindle count (total n) and density (spindles per minute) 191

Table 40. Summary of the key findings and impact of the congruency manipulation across Experiments 1-5 205
Table 41. Behavioural measures and their uncorrected relationship with sleep physiology in Experiments 2-5 209

List of Figures

Figure 1. Theoretical distinctions between memory subsystems. 16
Figure 2. An example of system-level consolidation between the hippocampus and neocortex. 19
Figure 3. The architecture of sleep across a normal night... 25
Figure 4. The active consolidation account of sleep. 31
Figure 5. The synaptic homeostasis hypothesis.. 34
Figure 6. A figure conceptually detailing some of the different transformative changes memories can undergo through consolidation. 41
Figure 7. The transitive inference paradigm with six elements... 51
Figure 8. The serial position effect and memory integration. 55
Figure 9. The semantic distance effect and memory integration. 57
Figure 10. Experiment 1 design... 58
Figure 11. Experiment 1: The adapted transitive inference paradigm. 59
Figure 12. Experiment 1: The congruency manipulation. 59
Figure 13. Experiment 1: The counterbalanced congruency conditions featured in training. 61
Figure 14. Experiment 1: Accuracy and RT performance between Delay type and Premise type. 67
Figure 15. Experiment 1: Delay type and Inference type RT (msec) and Training conformity (%) performance collapsed across Congruency. 70
Figure 16. Experiment 1: Congruent and Incongruent size judgment distances aggregated across Delay type. 72
Figure 17. Experiment 2 design... 87
Figure 18. Experiment 2: Training hierarchies. 88
Figure 19. Experiment 2: Method for sleep spindle detection. 92
Figure 20. Experiment 2: Accuracy and accurate reaction times for end and central items across all three experimental groups. 96
Figure 21. Experiment 2: Transitive Inferences across all testing blocks. 97
Figure 22. Examples of the A-B, A-C list learning paradigm. 118
Figure 23. Examples of how the relative distances between real words may influence the subsequent generalisation of novel words to new but familiar real word items........... 120
Figure 24. Experiment 3 design... 123
Figure 25. Experiment 3: The model 101 Zeo© home sleep recording device and wireless headband. 129
Figure 26. Experiment 3: 2AFC accuracy levels at test across Congruency, List type and Delay type. 136
Figure 27. Experiment 3: Generalisation levels (%) correlated with the differences in size focus between real word items for both congruent and incongruent responses across the 24 hour time delay. 139
Figure 28. Experiment 4: 2AFC accuracy levels at test across Congruency, List type and Delay type. 154
Figure 29. Experiment 4: Generalisation levels (%) correlated with the differences in size focus between real word items for both congruent and incongruent responses across the 24 hour time delay. 157
Figure 30. Experiment 5 design... 175
Figure 31. Experiment 5: An interaction between the levels of incorrect list recall between Delay type and List. 183
Figure 32. Experiment 5: 2AFC Accuracy levels at test across Congruency, List type and Delay type. 185
Figure 33. Experiment 5: Generalisation levels (%) correlated with the differences in size focus between real word items for both congruent and incongruent responses across the 24 hour time delay. 188

Figure 34. Diverging patterns of consolidation-associated integration seen in Experiment 1 207
Figure 35. The role of consolidation in the spontaneous recovery and stabilisation of congruent list memories over 24 hours. 208
Figure 36. The main behavioural findings of the present thesis. 214

Acknowledgements

First and foremost I would like to thank my supervisor Gareth Gaskell for being such an excellent and supportive mentor over the years. I am also grateful to the members of my thesis advisory panel, Beth Jeffries and Tom Hartley, who have always provided useful advice and insights into my research. My gratitude also extends to members of the sleep language and memory (SLAM) laboratory, both old and new, who have been friendly and engaging throughout my time here in York.

Many people have helped me in various ways throughout my PhD experience both in a professional and personal capacity. In fact, to thank everyone individually would necessitate an entire extra volume to this thesis. As no one wants that I hope it is enough to say that I am, and always will be, sincerely grateful to have such an amazing set of friends. A special thank you, however, must be extended to my parents: Sally and Gordon McGregor Reid. Without their continued and unconditional support this thesis simply would not have been written.

Dedication

I would like to dedicate this thesis to my grandmother Patricia Brown who sadly passed away before its completion. I'm not sure if she ever entirely understood what I was doing (despite my numerous attempts to explain) but I know she would be pleased that I saw it through to the end.

Declaration

This thesis contains original work completed solely by the author under the supervision of Professor Gareth Gaskell.

The research was supported by a studentship from the Biotechnology and Biological Sciences Research Council.

Data from this thesis was presented at the following conferences:

Experiment 1:

Reid, A.M., Gaskell, M.G. (2011) *Sleep and Semantic Memory Consolidation*. Poster presented at the Plasticity Across Systems conference, University of Lüebeck, Germany.

Experiment 1:

Reid, A.M., Gaskell, M.G. (2011) *Semantic Memory Consolidation over 24 Hours*. Poster presented at the International Conference of Memory 5 (ICOM 5), University of York, UK.

Experiments 3-5:

Reid, A.M., Gaskell, M.G. (2012) *The Role of Sleep and Semantic Memory in Retroactive Interference Recovery*, Talk given at the Sleep and Memory Symposium, Psychological Postgraduate Affairs Group conference, Northumbria University, UK.

Experiment 5:

Reid, A.M., Gaskell, M.G. (2012) *The Effects of Sleep and Time on Semantic Memory Stabilisation and Integration*. Poster presented at The European Sleep Research Society, Paris, France.

Chapter 1: Sleep, prior knowledge and dynamic memory consolidation

1.1 Introduction

Our remarkable ability to learn, retain and use information is often done so effortlessly that we take it for granted. Although this process may sometimes seem automatic, akin to uploading and downloading data from a hard drive, memories do not form literal records of past events. Instead, what we know comes from a series of interconnected associations that are continuously updated and reorganised in relation to our pre-existing knowledge. For example, memories may generalise to permit their retrieval in appropriate contexts, help link together useful elements of conceptual information and generally facilitate our predictive behaviour in relation to the world around us.

Nearly a century of research suggests that for memory to manifest these dynamic properties they must undergo a period of what is known as consolidation (Fowler, Sullivan, & Ekstrand, 1973; Gais, Molle, Helms, & Born, 2002; Jenkins & Dallenbach, 1924; McGaugh, 2000; Van Ormer, 1932). This process refers to a cascade of molecular, cellular and system-level processes that can bring about useful and enduring changes to long term memory (Dudai, 1996; 2004). Importantly, sleep appears to provide an optimal state for this phenomenon, permitting the ongoing stabilisation and enhancement of memory via neurological processes that might otherwise disrupt waking consciousness (Stickgold, 2005, 2009; Tononi & Cirelli, 2003, 2006). Indeed, a large body of research has causally implicated aspects of sleep physiology in the consolidation of memory (Rasch & Born, 2013).

While the above products of consolidation can relate to direct changes in memory performance, sleep has also been found to facilitate the reconstruction and reorganisation of memory. Behaviourally, this can manifest in the generalisation, abstraction and assimilation of new information in relation to what is already known (Inostroza & Born, 2013; Kroes & Fernández, 2012; Stickgold & Walker, 2013). Furthermore, integration in this manner may also help ‘salvage’ incomplete or weakly learned information (Drosopoulos, Schulze, Fischer, & Born, 2007; Fenn, Nusbaum, & Margoliash, 2003). These aspects of consolidation strongly implicate a top-down influence of prior knowledge in the ongoing refinement of new information. To date, however, both the behavioural outcomes and mechanisms underlying consolidation in this manner remain underexplored.

As such, the present thesis is driven by the overarching research question: how are new memories consolidated in relation to existing memory? This question can be segregated into two main themes: (i) *Compatibility*: how do memories that are contradictory or consistent with prior knowledge differ when it comes to consolidation?

And (ii) *Mechanisms*: what aspects of sleep physiology might be associated with these processes?

To develop these questions further the following literature review will discuss the relevant historical and theoretical contexts relating to memory and sleep in sections 1.2 and 1.3 respectively, while section 1.4 will focus more specifically on memory integration and transformation. Lastly, in section 1.5, the specific outline and research goals of the present thesis will be discussed.

1.2 Memory systems and consolidation

It has long been understood that memory is not a single monolithic construct but is best understood when meaningfully delineated into subsystems based on their functional and temporal qualities (Squire, 1992a). One of the most prominent distinctions is between declarative and non-declarative memory (Squire & Zola, 1996)(Figure 1). Declarative memory, as the name suggests, refers to knowledge and information about the world that can be consciously articulated (Eichenbaum, 2010) which itself can be divided into two further independent, yet related, subsystems: episodic and semantic memory. Episodic memory pertains to knowledge of events and experiences anchored within specific temporal or spatial context (Tulving, 1972; 1985, 2002). In contrast, semantic memory refers to a broader context-free form of knowledge, such as our conceptual understanding of objects, animals and events, and the relationships between them (Murphy, 2004; Rogers & McClelland, 2004). These systems are considered distinct from non-declarative memory[1] which relates to a collection of skills and habits typically acquired outside of conscious awareness (Squire, 1992a). It is worth noting that although these systems have often been described as distinct they likely operate in parallel in many instances of real world learning, such as language acquisition (Rasch & Born, 2013).

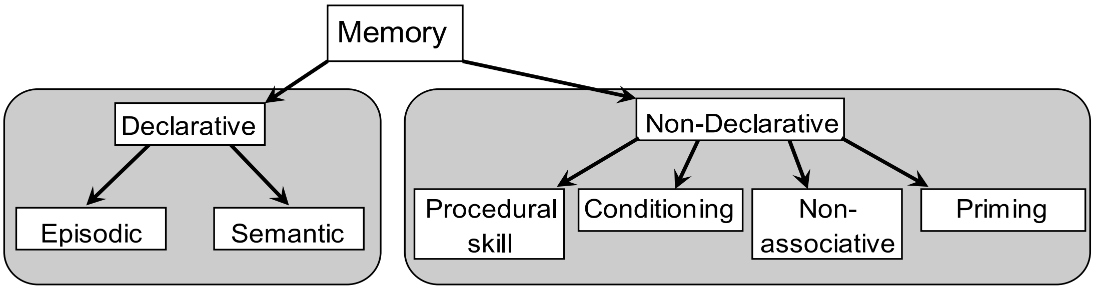


Figure 1. Theoretical distinctions between memory subsystems. This figure is adapted from Stickgold and Walker (2013).

---

[1] From this point onwards the terms 'non-declarative' and 'procedural' memory will be used interchangeably.

1.2.1 Consolidation

A modern understanding of memory encompasses three temporally related processes: encoding, consolidation and retrieval. While encoding relates to the initial acquisition of memories, and retrieval the subsequent access to this stored information, consolidation relates to the changes that stored memories can undergo between these two time points (Sara & Hars, 2006).

The observation that memory can change over time has a history extending back over the last century (Lechner, Squire, & Byrne, 1999; McGaugh, 2000). The term "consolidation", which itself is derived from the Latin term "to make firm", was coined through pioneering research by Muller & Pilzecker (1900) who examined the time course of learning by asking participants to memorise pairs of novel words. When memory was tested over a series of short time delays it was found that recall of these materials was vulnerable to disruption from exposure to additional novel words. However, when this time window was extended such disruptions ceased to have an effect, suggesting that this information had eventually become 'fixed' in memory. Independent research from the same era noted that amnesic patients with memory loss for past events, a phenomenon known as retrograde amnesia, manifested a temporally graded pattern of forgetting (Ribot, 1882). Specifically, the closer in time memories were formed to a traumatic amnesic event, the less likely they were to be preserved. Therefore, at the turn of the century two independent time courses for consolidation were observed: a short period over several minutes and a longer period extending days, months and potentially even years.

It was only until the mid-twentieth century that additional theoretical (Buxton, 1943) and experimental work (Duncan, 1945; 1949) began to further delineate these processes. An increased understanding of neuronal plasticity motivated Hebb (1949) to propose a potential mechanism for short term memory fixation based on the principle that neurones that 'fire together, wire together'. Reverbaratory mechanisms between brain cells extending beyond initial learning were argued to promote the ongoing stabilisation of memory. Later research by Marr (1970, 1971) would acknowledge the so called *stability-plasticity dilemma*: how can a memory system retain old memories while learning new ones?

As a solution Marr proposed that the twin demands placed on memory retention could be mapped onto structurally distinct areas of the brain that serve related, yet independent, functions. Initially acquired "simple memories", determined to reside in the archicortex (a subcomponent of the limbic system, including the hippocampus) were thought to be gradually transferred to the neocortex over time. This delegation in mnemonic processing was thought to capitalise on useful properties inherent in each subsystem, with the archiocortex permitting the rapid encoding and storage of new episodic information, while full integration into the overlapping semantic representations in the neocortex required longer periods of time.

1.2.2 The standard model of memory consolidation

The work of these pioneering researchers helped stimulate a concerted research effort into the time-course and mechanisms underlying consolidation over the subsequent decades, drawing from the fields of neuropharmacology and molecular genetics (Agranoff, Davis, & Brink, 1966; Kim & Fanselow, 1992; Nader, Schafe, & Doux, 2000). The theoretical groundwork initiated by Hebb (1949) and Marr (1970, 1971) are considered forerunners for two temporally distinct modes of consolidation known today as synaptic and systems consolidation respectively, both of which are integral components of *the standard consolidation model* (Dudai 1996; 2004).

Synaptic consolidation, which occurs within the first minutes or hours after encoding, refers to the local processing of experience dependent representations at a neural level. In parallel, or as a consequence of synaptic consolidation, systems-level consolidation relates to additional neuroanatomical shifts in memory storage that occur over 24 hours and beyond (Frankland & Bontempi, 2005). In terms of the previously established distinction between non-declarative and declarative memory, the former is thought to rely predominantly on synaptic consolidation (although see Dudai, 2004) while declarative memory is thought to rely on both (Squire, 1992a). Moreover, systems-level processes are thought to convert episodic memories into more decontextualised semantic forms over time (Inostroza & Born, 2013; McClelland & Rogers, 2003).

A prominent account relating to the principles of the standard consolidation model proposes a complementary learning system (CLS) between the hippocampus and neocortex (McClelland, McNaughton, & O’Reilly, 1995). The CLS implements the principles of connectionist modelling to predict the dynamic interplay of consolidation between brain systems (O’Reilly & Norman, 2002; O’Reilly & Rudy, 2001). Information is initially encoded in medial temporal lobe (MTL) structures and the neocortex in parallel. However, the hippocampal system, with its relatively high neuronal density and sparse interconnectivity, is thought to be specialised for the rapid acquisition of new information (Treves & Rolls, 1992) while more distributed neocortical structures are involved in the extrapolation and formation of statistical inferences about the world (McClelland & Rogers, 2003). As with the original proposal outlined by Marr (1970, 1971), these declarative subsystems are not insular, and interact to shift the retrieval dependency of episodic information to more integrated neocortical structures over time (Frankland & Bontempi, 2005, 2006) (Figure 2).

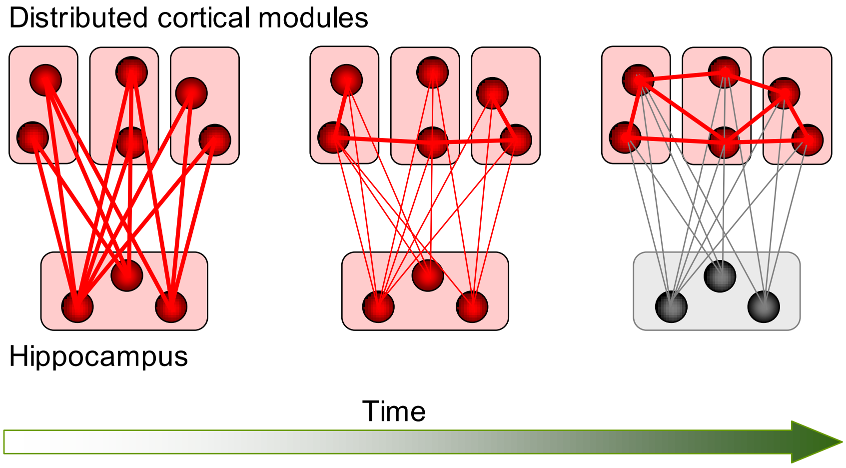


Figure 2. An example of system-level consolidation between the hippocampus and neocortex. Over time memories residing in the hippocampus are gradually integrated into the neocortex. Successive reactivation of this hippocampal-cortical network eventually allows an independent memory trace to form in the neocortex. The strength of connections in the hippocampal system are considered transient while connections in the neocortex are considered more robust and longer lasting. This figure is adapted from Frankland and Bontempi (2005).

In particular the connectionist principles underling the CLS outlined the way consolidation operates to offset the weaknesses, and maximise the strengths, inherent in its component subsystems. Computationally, a potential problem with the integrated neocortical system is that the introduction of new information, particularly if it is incompatible with what is already known, can rapidly disrupt the pre-existing connection weights in a computational dilemma called *catastrophic interference* (French, 1999). The CLS proposes to ameliorate this problem through the incremental transfer of new memories into the neocortex through successive, hippocampally mediated reactivations (Treves & Rolls, 1994). These reactivations serve to interleave new and existing information in a manner that allows an integrated and independent memory representation to form in the neocortex, while simultaneously unburdening hippocampal structures from their neural load (Frankland & Bontempi, 2005).

Importantly, the interaction between the memory systems driving this process is emphasised as being bi-directional, and new memories are not simply ‘shunted’ into long term memory via a hippocampal-neocortical route. Rather, dialogue between these systems permits the ongoing refinement and adjustment of what is already known in the face of new information (McClelland et al., 1995; Stickgold & Walker, 2013).

1.2.3 Evidence for the standard model

While synaptic consolidation has become established in its own right (see Pace-Schott & Hobson, 2002; Rasch & Born, 2013; Wixted, 2004) the rest of this section will be devoted to a short overview of evidence relating to systems-level consolidation given its relevance to the central research questions of this thesis. Although a full account of the literature is beyond the scope of this review, selected examples will be drawn from three areas representing the convergent evidence for this model: neuroanatomical studies of amnesiacs, the neurobiology of animal models and functional neuroimaging in humans.

Neuroanatomical studies of amnesiacs

A central line of evidence for the standard consolidation model relates to the dissociation between semantic and episodic memory systems in individuals with brain damage. It has repeatedly been observed that when MTL structures are compromised a 'classic' retrograde forgetting curve will emerge, as well as a degree of concurrent anterograde amnesia, producing simultaneous difficulties in acquiring new memories (Milner, 1972; Scoville & Milner, 1957). These findings strongly suggest that MTL structures are required for both the acquisition and post-encoding processing of new memories. Moreover, the extent of damage to the MTL is typically predictive of the severity of episodic memory deficits, with an expanded locus of damage to these structures manifesting more severe and extended forms of graded amnesia (MacKinnon & Squire, 1989; Squire, Haist, & Shimamura, 1989). Conversely, individuals with damage restricted largely to the neocortex and not the hippocampus, such as patients with semantic dementia, tend to manifest a progressive loss of conceptual knowledge while episodic memories remain comparatively intact (Graham, Becker, & Hodges, 1997; Snowden, Griffiths, & Neary, 1996).

While these findings confirm the double dissociation between the brain regions and predicted temporal dynamics outlined in the standard consolidation model it should be also be noted that research using amnesic populations is not without certain drawbacks. These studies are necessarily opportunistic in nature, relying on individuals who have acquired brain damage in ways that do not necessarily conform with the precise regions of interest to the researcher (Meeter & Murre, 2004). Furthermore, when 'pure' cases are found they are often small in number and may manifest variable patterns of forgetting depending on the type of methodology used to probe memory (Nadel & Moscovitch, 1997). Therefore, although amnesic studies provide a solid foundation of evidence for the neuroanatomical distinctions made in the standard model of memory consolidation, other lines of evidence are necessary to fully corroborate its predictions.

Animal neurobiology

A benefit of using animal models is that they permit the systematic manipulation of analogous neural substrates implicated in the standard consolidation model. Investigations using both rats (Bontempi, Laurent-Demir, Destrade, & Jaffard, 1999) and monkeys (Zola-Morgan & Squire, 1990) have so far confirmed the time dependent shift in retrieval dependency from hippocampal to neocortical structures. However, these investigations tend to focus exclusively on the feed forward transfer of spatial or episodic-like memories into the neocortex independent of their relevance to prior knowledge.

The few experiments exploring the latter aspect of the systems-level dynamic suggest that prior-knowledge forms an important gating mechanism for the uptake of new information. For example, Tse et al. (2007) manipulated the compatibility of trained information in relation to prior knowledge by exposing rats to a sand filled arena with fixed food locations following a predictable 'schematic' layout cued with a series of smells (e.g. 'banana'). Hippocampal or sham (control) lesions were administered at different points during in training, with learning measured through appropriate digging at reward sites. Interestingly, rats lesioned after 48 hours could generalise to novel but schematically consistent reward probes, whereas rats lesioned after 3 hours could not, suggesting that a shift in retrieval dependency over time is required for the generalisation of new information. Moreover, when the training of food-location schemas was administered in a random (inconsistent) or predictable (consistent) manner it was found that inconsistent training blocked both task improvement and the subsequent capacity for generalisation. These findings not only validate the existence of system-level mechanisms, but also illustrate how the compatibility of new information with pre-existing knowledge can potentially expedite or block subsequent neural integration.

Neuroimaging studies in humans

Lastly, neuroimaging techniques in human participants, such as functional magnetic resonance imaging (fMRI), provide a third confirmatory line of support for the standard consolidation model. These findings are typified in a series of experiments conducted by Sterpenich and colleagues (2007;2009) who monitored the neural activation associated with the recall of pictures three days, and six months, after initial exposure. Compared to the initial recall session, delayed recall produced a large shift in activation away from hippocampal structures to various cortical locations. Similarly, a study by Takashima et al. (2009) explored the neural correlates of time delayed memory in a face-location task. Items tested 24 hours after learning, as opposed to 15 minutes, produced an activation profile with diminished hippocampal involvement, but increased functional connectivity between task-relevant neocortical structures. Interestingly, the particularly rapid consolidation observed in this latter study may have been accelerated, at least in part, by previous exposure to a similar but untested face-location task administered a week prior. As with the observations made by Tse et al. (2007) these findings potentially implicate relevant prior knowledge in the accelerated integration of new information.

Despite the neuroanatomical evidence relating to the time course of systems-level consolidation in humans the influence of prior knowledge, as with animal research, remains a relatively underexplored component of the standard model. Indeed, the majority of the paradigms featured in the literature typically invoke hippocampally dependent memory tasks with little or no relationship to semantic memory (van Kesteren, Ruiter, Fernández, & Henson, 2012). What evidence there is suggests that learning tasks that contradict pre-existing associative networks can induce enhanced hippocampal-neocortical connectivity after a period of consolidation, implicating potential compensatory processes, while compatible information tends to manifest the expected (and rapid) shift in retrieval dependency into neocortical structures (van Kesteren, Fernández, Norris, & Hermans, 2010). These preliminary findings further confirm that the transfer of new memories from MTL structures to the neocortex is not a purely bottom-up process, and that prefrontal structures may exert top-down influence when integrating new information (Inostroza & Born, 2013; Kroes & Fernández, 2012).

1.2.4 Multiple Trace Theory

It should be noted that the standard model has not remained uncontested. It has been argued by some that this account conflates the potentially independent processing of semantic and episodic information (Moscovitch & Nadel, 1998; Nadel & Moscovitch, 1997). Furthermore, amnesic patients often display retrograde memory deficits extending decades, argued to be a biologically implausible length of time required for consolidation (Nadel, Winocur, Ryan, & Moscovitch, 2007). To this end, Multiple Trace Theory (MTT) proposes that rich contextual details relating to episodic and autobiographical memory rely on multiple memory traces generated in the hippocampus that remain linked to corresponding neocortical networks indefinitely. However, as semantic memory is considered context free, consolidation is predicted to occur in the same manner as outlined by the standard consolidation model. Consequently, a testable prediction of the MTT is that extensive damage to the hippocampus should produce larger detrimental effects for remote episodic, but not semantic memories (Moscovitch, Nadel, Winocur, Gilboa, & Rosenbaum, 2006).

Evidence supporting this view includes cases where amnesic patients exhibit flat retrograde forgetting curves after hippocampal damage (Cipolotti et al., 2001; Rosenbaum, Winocur, & Moscovitch, 2001) as well ongoing involvement of the hippocampus in retrieving remote memories formed up to 35 years prior (Maguire, Henson, Mummery, & Frith, 2001; Ryan et al., 2001). Additionally, there is evidence for distinct post-encoding time courses for the consolidation of semantic and episodic memory (Barr, Goldberg, Wasserstein, & Novelly, 1990). This is exemplified in a case study by Warrington and McCarthy (1988) that examined the vocabulary acquired by an individual during a period of dense retrograde amnesia. In contrast to his profound impairment of episodic information, the patient displayed a striking retention for words that had entered his vocabulary during in the amnesic period, such as ‘shuttle’, ‘AIDS’ and ‘thatcherism’. This study, along with others (Moscovitch et al., 2006; Verfaellie, Reiss, & Roth, 1995), implicate dissociable consolidation processes for episodic and semantic systems.

Although the case for MTT is coherent it nevertheless has its own set of discrepancies. If MTT is treated as falsifiable on the basis that remote episodic memories can arise without hippocampal involvement then the model quickly encounters problems. To this end, a number of studies have observed the ongoing reduction in hippocampal activity at remote time points (Bontempi et al., 1999; Frankland, O’Brien, Ohno, Kirkwood, & Silva, 2001; Teng & Squire, 1999), including the previously discussed neuroimaging investigations (Sterpenich et al., 2009; Takashima et al., 2009). Furthermore, in patients exhibiting flat retrograde forgetting curves for episodic details, seemingly at odds with the predictions made by the standard model, the locus of brain damage has often been found to include neocortical areas, such as the anterior and lateral temporal lobes (Farah & McClelland, 1991; Zola-Morgan & Squire, 1993). Lastly, it has been noted elsewhere that criticisms relating to the unfeasible time course required for consolidation can actually be applied to MTT itself, as both models are in agreement on the existence of systems-level consolidation for semantic memory (Meeter & Murre, 2004).

Many of the issues raised by MTT remain the topic of ongoing debate (see Inostroza & Born, 2013). Nevertheless, the present consensus in the literature appears to weigh in favour of the standard consolidation model, which is generally considered the more established of the two accounts (Dudai, 2004). That said, proponents of MTT rightly pose questions about some of the conflicting observations in the literature, and future researchers should by no means treat the textbook account of consolidation as immutable.

1.2.5 Conclusions

It should be clear from this section that there is a strong case for systems-level consolidation. Nevertheless, despite the multiple lines of evidence several issues remain. Firstly, the time course for consolidation is somewhat erratic depending on the methodology involved. Amnesic patients implicate systems-level processes spanning years (MacKinnon & Squire, 1989; Ryan et al., 2001), animal models suggest days and weeks (Bontempi et al., 1999; Tse et al., 2007) while fMRI data indicate that rapid consolidation can occur in just one day (Takashima et al., 2009). Indeed, the theoretical work underpinning the standard consolidation model itself remains somewhat open-ended on the precise time period required to consolidate new information (Meeter & Murre, 2004).

One possibility is consolidation involves multiple waves of stabilisation and integration, leading to progressive memory refinement over time (Cairney, 2012; Rasch &

Born, 2013). This interesting prospect also relates to the noted paucity of experimental research exploring how newly consolidated information might interact with pre-existing knowledge. Manipulating the relationship trained information has with existing semantic concepts may help tease apart factors relating to the nature and speed of integratory processing. As it stands information that is congruent with prior knowledge appears to be particularly amenable to a rapid 'fast tracked' time course for systems-level integration (Takashima et al., 2009). Conversely, the uptake of incongruent information, depending on its applicability to future behaviour, may remain reliant on hippocampal structures, and is perhaps even actively blocked from higher order integration altogether (Tse et al., 2007; van Kesteren, Fernández, et al., 2010).

Despite the ongoing debate relating to the consolidation of declarative information, it is agreed by both the multiple trace and standard consolidation accounts that systems-level changes to memory can occur over time. This prospect itself raises interesting additional questions. For example, when exactly do these processes occur? What are the physiological correlates of integration? The answer to these questions, as shall be elaborated, relates to sleep and its influence on memory consolidation.

1.3 Sleep and memory consolidation

Sleep is defined as 'a natural periodic state of rest during which consciousness of the world is suspended' (Spriggs, 2008, p. 10) and is a behaviour observed in all humans and animal phyla observed thus far (Cirelli & Tononi, 2008). We spend around one third of our lives asleep, and when this state is deprived or truncated serious psychological and physiological ramifications can occur (Montagna & Lugaresi, 2002). Although the biological drive for sleep is undeniable a single core function has yet to be found, with the current findings from biological, neurological and behavioural domains indicating a collection of roles centred around homeostasis, restoration and memory consolidation (Aldrich, 1999).

1.3.1 Sleep characteristics

Historically speaking, sleep has often been misconstrued as a catatonic state with little or no importance beyond the need for rest itself (see Aldrich, 1999). Contrary to this notion sleep is in fact a highly active state and, like memory, is not a homogenous construct. Sleep can be meaningfully dissociated into a number of distinct sub-stages based on their respective electrophysiology (Rechtschaffen & Kales, 1968), neurochemistry (Pace-Schott & Hobson, 2002), homeostatic regulation (Borbély & Achermann, 1999) and regional brain activity (Maquet et al., 1996). Consequently, the architecture of the sleep state can be understood on multiple levels ranging from its component sub-stages to the lower level intra-state elements spanning the entire sleep cycle (Conte & Ficca, 2012). Before discussing how the composition of sleep relates to memory consolidation several of these fundamental distinctions must be clarified.

In terms of sleep stages, a primary distinction can be made between rapid eye movement (REM) and non-REM (NREM) sleep. The former is defined by relative muscle atonia, frequent bursts of eye motility and electrical activity in the brain similar to that of wakefulness (Aserinsky & Kleitman, 1953). NREM sleep, on the other hand, refers to a collection of additional sub-stages, ranging from stages one to four, which reflect the ascending 'depth' of sleep (Rechtschaffen & Kales, 1968). Over the course of an average night of sleep REM and NREM stages tend to flow into each other over successive 90-minute cycles (Hobson & Pace-Schott, 2002). Due to a combination of homeostatic and circadian pressures NREM typically predominates over the first half of the night, and REM the latter half, respectively (Walker & Stickgold, 2006) (Figure 3, A).

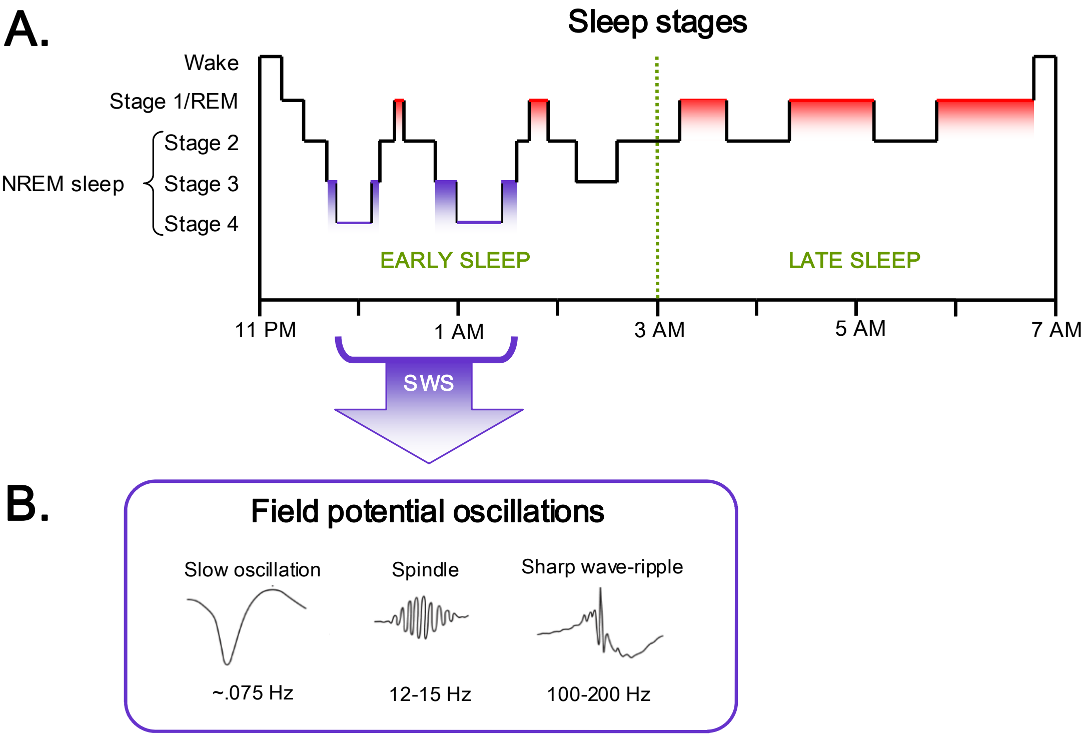


Figure 3. The architecture of sleep across a normal night. (A) Sleep stage activity. Red bars indicate periods of rapid eye movement (REM) sleep while the deepest stages of sleep (purple shaded stages 3 and 4) are collectively known as slow wave sleep (SWS). NREM sleep denotes non-rapid eye movement sleep (stages 2-4). (B) Field potential oscillations and frequencies associated with SWS as measured through electroencephalography (EEG). The images of field potential oscillations in this figure are taken from Rasch and Born (2013).

Additional distinctions relating to sleep physiology can be made on the basis of the presence, or absence, of a number of low level field potential oscillations spanning the sleep stages defined above (Steriade, 2006). In terms of declarative memory consolidation the three most important of these field potentials are: (i) slow oscillations (~.075 Hz), (ii) sleep spindles (13-15 Hz), and (iii) sharp wave ripples (100-200 Hz) which are defined in more detail below (Figure 3, B) (Silber et al., 2007; Steriade, 2006). These oscillations, which are typically registered through electroencephalography (EEG), are found in the deepest stages of sleep (stages three and four) collectively known as slow wave sleep (SWS). The dynamic involvement of both sleep stage activity and these low level field potentials in memory processing will be discussed over the following sections.

1.3.2 Sleep dependent memory consolidation

The potential benefits of sleep on memory function have also been acknowledged by academics and philosophers for some time (Hartley, 1834; Quintillan, 95AD) including Ebbinghaus (1885) who noted that declarative memories appear to be more sustained by sleep than daytime wakefulness. Modern empirical work has framed these early suppositions into four testable hypotheses (Ellenbogen, Payne, & Stickgold, 2006): (i) that sleep confers no advantages to memory; (ii) that sleep *does* confer an advantage, but only though protecting memories from interference; (iii) that sleep provides a passive-protective environment for consolidation processes to take place, and (iv) that sleep confers a unique biological state that permits the active consolidation of information. These positions encompass a long standing debate in the literature: does the sleep state provide passive protection (ii, iii) or an active contribution (iv) to the consolidation of memory?

1.3.3 Models relating to sleep stage activity

The first empirical attempt at measuring the influence of sleep on memory was conducted by Jenkins and Dallenbach (1924) who compared the forgetting rate of nonsense syllables across periods of sleep and wakefulness using a series of free-recall sessions. It was found that periods of learning punctuated by sleep were marked by a relative increase in the retention of these learned materials. Jenkins and Dallenbach themselves speculated that these benefits to memory were passive-protective in nature: ‘The results of our study as a whole indicate that forgetting is not so much a matter of the decay of old impressions and associations as it is a matter of the interference, inhibition, or obliteration of the old by the new.’ (Jenkins & Dallenbach, 1924, p. 612). However a later and more sophisticated replication of the Jenkins and Dallenbach paradigm demonstrated a more active role of sleep in consolidation, as stabilisation effects were still observed once waking interference was controlled for (Benson & Feinberg, 1977). Research into this area has enjoyed a prominent revival from the 1970s onwards, with research paradigms typically probing two mechanistically distinct forms of consolidation: increased resistance to interference (‘stabilisation’), and improvements in veridical memory performance (‘enhancement’) (Walker, 2005).

A solid body of evidence has since accumulated regarding the stabilising effects of sleep for both procedural (Korman et al., 2007) and declarative memories (Ellenbogen, Hulbert, Jiang, & Stickgold, 2009; Ellenbogen, Hulbert, Stickgold, Dinges, & Thompson-Schill, 2006). While stabilisation can also be observed over wakefulness (Brashers-Krug, Shadmehr, & Bizzi, 1996) benefits in terms of enhancement appear to derive primarily from sleep in both the declarative (Plihal & Born, 1997; Tucker et al., 2006) and procedural domains (Fischer, Hallschmid, Elsner, & Born, 2002; Walker, Brakefield, Morgan, Hobson, & Stickgold, 2002). For example, a nap as short as six minutes can bolster memory for declarative information relative to wakefulness (Lahl, Wispel, Willigens, & Pietrowsky, 2008) while a 3-hour post-training nap can enhance the retention for materials up to four years later (Wagner, Hallschmid, Rasch, & Born, 2006).

The sheer range and proliferation of research into this area over the past few decades produced two schools of thought regarding sleep stage associated memory processing: *task-dependent models*, postulating that different mnemonic processes depend on the underlying sleep stages involved, and *sequential models*, that consider consolidation operating as a function of the unified co-operation between sleep stages (Conte & Ficca, 2012).

Task dependent models

The task-dependent approach initially derived from theoretical accounts proposing REM sleep as an active mediator in consolidation (Empson & Clarke, 1970). One experiment exploring this possibility measured how the disruption of REM sleep might influence procedural memory performance (Karni, Tanne, Rubenstein, Askenasy, & Sagi, 1994). Participants were trained on a simple visual discrimination task followed by either a normal night of sleep, or sleep with disruptions specific to the REM or SWS stages. When memory was probed after a sufficient recovery period it was found that selective REM stage disruption had exclusively blocked improvements observed in the other two conditions. It was concluded that REM sleep physiology was promoting the synaptic consolidation of recent perceptual experiences (Karni et al., 1994).

However, REM sleep disruption has produced mixed results in the declarative domain. While some studies noted deleterious effects on the ability to verbally recall sentences and stories (Empson & Clarke, 1970; Tilley & Empson, 1978) other investigations observed no such effect in paired associative word learning (Ekstrand, Sullivan, Parker, & West, 1971) or word fluency tasks (Lewin & Glaubman, 1975). Furthermore, the deprivation manipulation itself was criticised on the basis that any resulting changes in memory performance might be confounded by non-specific motivational factors, stress and cognitive deficits (Born & Gais, 2000; Cipolli, 1995). Nevertheless, the task-dependent position led to the conception of the *dual process hypotheses* which proposed that the differing physiology underlying REM and SWS is specialised for synaptic and systems-level consolidation respectively, thus conferring independent advantages for the processing of procedural and declarative memories.

Evidence for this model was established by Plihal & Born (1997) who avoided the methodical pitfalls inherent in deprivation techniques by taking advantage of the natural predominance of SWS (early) or REM (late) rich periods in nocturnal sleep (Yaroush, Sullivan, & Ekstrand, 1971)(Figure 3, A). Participants had their memory probed on declarative (paired associate) and procedural (mirror tracing) tasks after sleep onset timed to coincide with REM or SWS-rich sleep episodes. It was found that earlier SWS-rich sleep conferred a ~15% enhancement for declarative memories not seen in the later sleep group or a waking control condition. Conversely, earlier REM-rich sleep produced a ~20% increase in procedural memory performance, with no such improvements after SWS rich sleep or wakefulness. This double dissociation was taken as evidence that REM and SWS conferred distinct functional advantages for procedural and declarative memories respectively. A follow up investigation confirmed that this sleep stage dichotomy was dissociable on the basis of the memory systems involved and not the verbal/nonverbal nature of the learned materials (Plihal & Born, 1999).

However, problems with the dual process account can be raised on both methodological and theoretical grounds. Firstly, much of its supporting evidence derives from the early/late paradigm in which retrieval occurs at different points in the day, a factor that may be confounded by circadian influences (Christianson & Christianson, 1992). A similar issue can be raised regarding the lack of prior sleep in the ‘late’ condition (Grosvenor & Lack, 1984). Furthermore, the evidence points to a significantly more complex picture of consolidation than this binary model proposes (Schabus, 2009). Conflicting findings have weakened the dual process position in this regard, with spatial (Wamsley, Tucker, Payne, & Stickgold, 2010), and procedural motor tasks (Walker, Brakefield, et al., 2002) having been shown to be enhanced by NREM sleep. Furthermore, some studies have found that REM sleep activity is associated with the consolidation of both neutral (Fogel, Smith, & Cote, 2007) and emotional (Wagner, Gais, & Born, 2001) declarative memories.

Task independent models

Sequential models propose a more inclusive role for consolidation linked to the cyclic alternation of REM and NREM stages across sleep (Ambrosini & Giuditta, 2001). Much in the way that various organs in our digestive system operate together to process food, sleep stages are proposed to work cooperatively, and sequentially, to process memories (Cipolli, 1995; Giuditta et al., 1995). This account is perhaps intuitive given that, in terms of real world learning, procedural and declarative memories are often acquired in simultaneous and complementary fashion (Rasch & Born, 2013).

Supporting evidence for this model includes a study by Gais, Plihal, Wagner, and Born (2000) that examined the inclusive contributions of sleep stages on a visual texture discrimination task. It was found that discrimination skills were faster after a period of SWS-rich sleep, with no discernable improvements after REM-rich sleep or equivalent periods of wakefulness. However, performance after a full night of sleep, including both REM and NREM stages, conferred enhancements almost three times greater than that observed for SWS alone. Elsewhere, improvements in a similar task were found to strongly correlate with the amount of SWS obtained in the first quarter of a night of sleep, and REM sleep in the final quarter (Stickgold, James, & Hobson, 2000). Furthermore, verbal retention of learned material has been positively associated with the cohesion of prior NREM/REM cycles for both young adults (Ficca, Lombardo, Rossi, & Salzarulo, 2000) and elderly individuals (Mazzoni et al., 1999).

A primary concern with the sequential hypothesis, however, is testability. Studies tend to rely heavily on correlations, or extremely complex manipulations of sleep fragmentation, leading to a relative lack of evidence in human participants (Conte & Ficca, 2012). Additionally, the model lacks explanatory power for some of the more complex interactions observed between memory systems during sleep. For example, a study by Brown & Robertson (2007) manipulated the order of training for declarative (word learning) or procedural (serial reaction time) tasks prior to sleep onset. When declarative learning followed a procedural task subsequent declarative enhancement was blocked after a night of sleep, with the opposite pattern emerging when this task order was reversed. Participants who remained awake between training and testing, however, displayed no such interference effects.

This competitive dissociation suggests a complex interplay memory systems during sleep, particularly given that procedural memories have been found to facilitate the post-sleep enhancement of declarative memories elsewhere (Fischer, Drosopoulos, Tsen, & Born, 2006; Wagner, Gais, Haider, Verleger, & Born, 2004). Together these studies potentially implicate successive periods of cooperation and disengagement between declarative and non-declarative memory systems across the sleep cycle (Robertson, 2009). A purely sequential model, with its emphasis on feed-forward cooperation, therefore potentially neglects these important factors relating to the dynamic interaction between memory systems.

1.3.4 Models relating to field potential oscillations

The literature reviewed so far indicates that sleep can both enhance and stabilise procedural and declarative memories, yet there has evidently been some controversy regarding the precise role that sleep stages, and indeed the sleep cycle as a whole, have in this process. Potential resolutions to some of these issues may be derived from a complementary examination into the more fine-grained aspects of sleep physiology underlying these stages (Diekelmann & Born, 2010a; Stickgold, 2013).

Two prominent models have emerged from the literature as a result: *the active system consolidation hypothesis* (Born, Rasch, & Gais, 2006; Born & Wilhelm, 2012; Frankland & Bontempi, 2005) and *the sleep homeostasis hypothesis* (Tononi & Cirelli, 2003, 2006). Both of these models implicate slow wave sleep in the ongoing refinement of memory, although the precise mechanisms driving these processes differ considerably between the two accounts.

Active consolidation

Drawing on influences from the previously discussed experimental (Plihal & Born, 1997, 1999) and theoretical literature (Buzsáki, 1989; Marr, 1971; McClelland et al., 1995) the active consolidation model proposes a set of endogenous, sleep-based mechanisms that provide an elegant account of the standard model of memory consolidation presented originally in section 1.2.2. This account focuses on the synergistic interaction between the three field potential oscillations outlined previously in Figure 3, namely slow wave oscillations, spindles and sharp wave ripples (Rasch & Born, 2013).

Here, slow oscillations generated in the neocortex are thought synchronise the activation of widespread neuronal populations by globally inducing periods of neuronal hyperpolarisation (silence) or depolarisation (excitation) (Steriade, 2006). This synchronous activity operates in a feed-forward manner to establish a temporal framework for dialogue between the neocortex and relevant subcortical structures. Specifically, memory representations in the hippocampus are reactivated through high frequency bursts of sharp wave ripple activity, which in turn coalesce with spindle activity generated in the thalamus (Siapas & Wilson, 1998). The driving influence of the depolarising up-phase of slow oscillations on spindle and hippocampal ripple activity enables ripples to become nested into the single oscillatory troughs of spindles (Mölle, Eschenko, Gais, Sara, & Born, 2009; Mölle, Marshall, Gais, & Born, 2002).

As illustrated in Figure 4, these integrated spindle-ripple events operate in synchrony with the depolarising slow wave upstate to facilitate hippocampal-neocortical feedback at a point when neuronal networks are optimally predisposed to persisting synaptic changes (King, Henze, Leinekugel, & Buzsáki, 1999; Rosanova & Ulrich, 2005). The eventual product of this iterative process is the transfer and integration of initially hippocampal dependent memory traces into the neocortex.

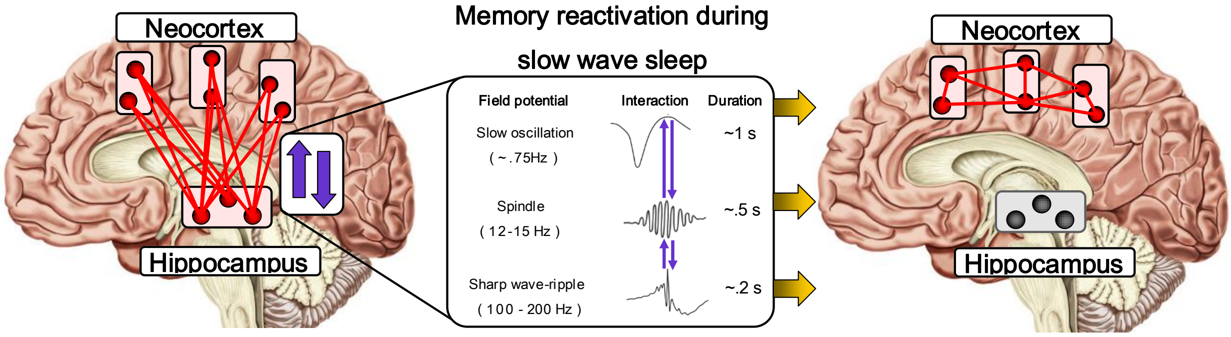


Figure 4. The active consolidation account of sleep. Purple arrows represent the bi-directional flow of information between the hippocampus and neocortex as guided by the conjunctive activation of slow oscillations, spindles and sharp wave ripples. Yellow arrows represent the products of this iterative process upon memory: an independent, integrated and potentially distributed memory trace in the neocortex. The brain in this figure is an illustration created by Giuseppe Polegri and used with permission (http://www.polegri.it/en/illustration.htm).

Evidence for active consolidation

Slow wave activity

A role for slow wave oscillations in declarative memory consolidation was established in a study by Marshall and colleagues (2006) who implemented transcranial direct current stimulation (tDCS) to induce slow (0.75 Hz) and fast (5 Hz) oscillating field potentials over the pre-frontal cortex during NREM sleep. As the 0.75 Hz frequency range is a natural feature of slow oscillations it was predicted that this manipulation would the stimulate feed-forward activation and enhancement of declarative memories. This supposition was validated, with 0.75 Hz stimulation simultaneously promoting slow oscillations, spindle activity and the enhanced retention of declarative materials. Procedural task performance, however, remained unaffected (Marshall, Helgadottir, Molle, & Born, 2006). More recently, the converse pattern was demonstrated through tDSC, with the application of a 5 Hz theta frequency during NREM sleep suppressing both slow wave oscillations and subsequent waking performance on a word-pair retention task (Marshall, Kirov, Brade, Mölle, & Born, 2011). Together these findings strongly implicate a causal role for SWS activity in the active consolidation of declarative memories.

Spindle activity

Sleep spindles have become established as an important bio-marker of memory consolidation. Spindle activity appears to occur preferentially at synapses directly involved with encoding (Werk, Harbour, & Chapman, 2005) and can predict improvements in performance on a range of declarative and procedural tasks (Clemens, Fabó, & Halász, 2005; Nishida & Walker, 2007; van der Helm, Gujar, Nishida, & Walker, 2011; Walker, Brakefield, et al., 2002). More recently, a causal role for spindles in memory consolidation has been implicated through the use of pharmacological interventions that stimulate their expression during sleep. Specifically, administration of the drug zolpidem was found to boost spindle activity and declarative memory retention in medicated individuals relative to a non-medicated control condition (Mednick et al., 2013).

While spindles occur throughout the NREM state, particularly during stage 2 sleep, their composition in SWS sleep may be of particular importance to declarative memory consolidation (Mölle et al., 2009). This notion was recently confirmed by Cox and colleagues who asked participants to attend to a short six minute film clip, with the subsequent recall of film elements tested both before and after a night of sleep. Polysomnographic EEG recordings indicated a >60% increase in SWS spindle density with no such change in NREM stage 2 sleep. Moreover, SWS spindles were strongly correlated with retention of the perceptual and conceptual details of the film (Cox, Hofman, & Talamini, 2012).

Lastly, the sleep spindle frequency range (12-15 Hz) has been found to comprise of two functionally distinct spindle types: fast spindles (<13 Hz), localised in anterior areas of the scalp, and slow spindles (13> Hz), displaying a more posterior distribution (Anderer et al., 2001). Fast spindles in particular have recently been proposed to influence consolidation by facilitating the occurrence of slow-oscillation-spindle cycles (Mölle, Bergmann, Marshall, & Born, 2011). While evidence has sometimes been mixed (Schabus et al., 2008) fast spindles have nevertheless been found to predict both performance enhancements and learning ability in the declarative domain (Saletin, Goldstein, & Walker, 2011; van der Helm et al., 2011) and may therefore be particularly integral to systems-level consolidation during sleep (Inostroza & Born, 2013).

Memory reactivation

Although fMRI investigations relating to memory consolidation have consistently uncovered neurological shifts in retrieval dependency after sleep (as noted in section 1.2.3) the bulk of the research discussed thus far has nevertheless relied on correlations between memory and sleep activity to infer active consolidation. An important line of evidence for the active consolidation account therefore relates to the mechanisms associated with mnemonic reactivation underlying the model.

To this end, recordings of hippocampal place-cells in rats have been found to manifest patterns of re-activation during SWS for neurones previously active during waking spatial learning (Pavlides & Winson, 1989; Wilson & McNaughton, 1994). Furthermore, the task-dependent content of these reactivations can be biased by re-presenting auditory cues originally present at training during subsequent SWS (Bendor & Wilson, 2012).

In humans the cueing of task relevant sensory information during sleep has been strongly associated with consolidation (Rudoy, Voss, Westerberg, & Paller, 2009). For example, an innovative experiment conducted by Rasch, Büchel, Gais, and Born (2007) exposed participants to a salient odour while they learned a series of card-object locations. A subset of these participants were then re-exposed to this odour during an ensuing sleep episode, with the aim of associatively triggering latent memory reactivation. As predicted, odour presentations in SWS conferred enhancements to memory not observed for presentations during periods of REM sleep or wakefulness. Interestingly, supplementary fMRI analysis indicated that SWS presentations promoted hippocampal activation to an even greater extent than prior waking exposure, suggesting the selective sensitivity of these structures to inputs capable of reactivating memories.

Memory selectivity

Lastly, the issue of memory selectivity must be raised as the active consolidation model necessitates some form of discrimination during encoding to avoid saturating neocortical structures with superfluous information (Born & Wilhelm, 2012). However, with research efforts to date focussing primarily on changes in post-consolidation performance this important factor arguably remains underexplored in the literature (Conte & Ficca, 2012). Nevertheless, at present, evidence suggests a number of psychologically mediated factors may ‘tag’ memories for subsequent consolidation.

It has been found that the explicit (aware), as opposed to implicit (unaware), nature of a trained task can facilitate post-sleep memory enhancement in the procedural domain (Song, 2009; Spencer, Sunm, & Ivry, 2006). Indeed, in some cases task performance improvements can be switched from sleep independent to sleep dependent simply by informing participants of underlying response patterns within a task (Robertson, Pascual-Leone, & Press, 2004). Furthermore, an fMRI study exploring emotional memory retention observed a selective enhancement of functional campo-neocortical connectivity during the recall of negative, but not neutral, stimuli (Sterpenich et al., 2007). During encoding the concurrent activation of the hippocampus and amygdalar system, a region associated with emotional processing, were thought to have facilitated the subsequent shift in retrieval dependency away from these structures during sleep.

Task salience, and more broadly the attentiveness of the participant, may therefore activate MTL structures during encoding (Diekelmann, Wilhelm, & Born, 2009) implicating their involvement as a prerequisite for subsequent consolidation (Saletin & Walker, 2012). Moreover, as noted in section 1.1, the top-down influence of prior-knowledge may have some role to play in the ‘tagging’ of new memories for consolidation depending on their complementary, or contradictory, relationship with pre-existing semantic networks (Kroes & Fernández, 2012).

The synaptic homeostasis hypothesis

In a second account of sleep-associated consolidation, the synaptic homeostasis hypothesis, proposes that waking interactions with the environment will amass an increasing net gain in synaptic strength across the brain, placing demands on both the volume of neural tissue and energy available. This potentially maladaptive scenario is thought to be ameliorated through the proportionate reduction of synaptic connectivity during sleep.

More specifically, the successive depolarisation and hyperpolarisation of slow wave oscillations, combined with a neuromodulatory milieu conductive for synaptic depression during SWS, are thought to promote a global downscaling of synaptic strength. This self-limiting mechanism, which is illustrated in Figure 5, is proposed to reduce, or even remove, weakly encoded information leaving synaptic connections strong enough to survive this process with a relatively enhanced signal to noise ratio (SNR). The model therefore not only accounts for sleep-associated improvements in memory performance, but also the drive for sleep and its underlying architecture, such as the early predominance of SWS across the night (Tononi & Cirelli, 2003, 2006).

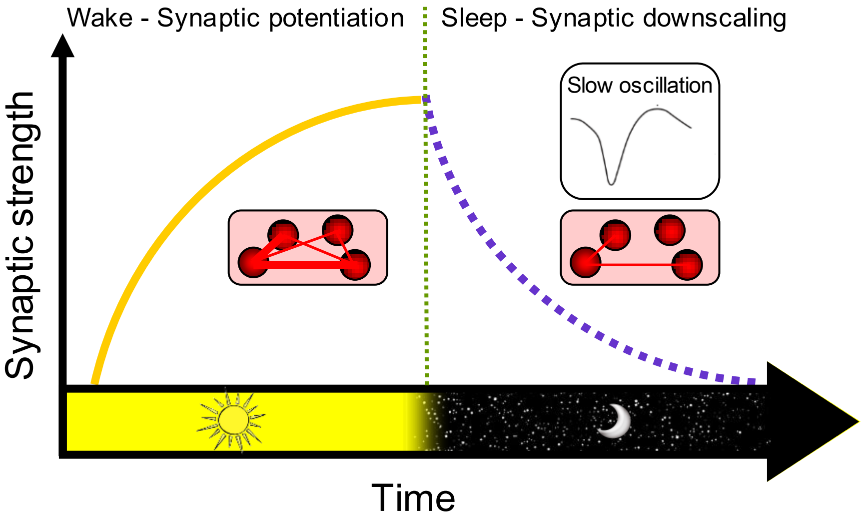


Figure 5. The synaptic homeostasis hypothesis. Synaptic potentiation is accrued across wakefulness (the yellow curve) to levels of varying strength between neurones (the thicker and thinner red lines). During sleep, slow oscillations globally downscale the synaptic connections (the purple dashed curve) leaving only the most robust associations. This figure is based on principles outlined by Tononi and Cerelli (2006).

Evidence for synaptic homeostasis

The synaptic homeostasis hypothesis predicts neurological activity that promotes long term synaptic potentiation (LTP) across wakefulness and concomitant long term depression (LTD) processes during SWS. Evidence relating to the former component of this model includes the observation that experimentally induced whisker stimulation during wakefulness can induce marked increases in synaptic density at the associated locus of cortical activity (Knott, Quairiaux, Genoud, & Welker, 2002).

With regards to the latter component of the model, LTP appears to be reduced in sleep through the diminished expression of plasticity-related genes during SWS (Cirelli & Tononi, 2000) suggesting a dichotomous neurological milieu across wakefulness and sleep favourable for encoding and synaptic downscaling respectively. Furthermore, as previously noted, the increasing pressure for depotentiation accrued over the course of wakefulness may drive the increasing need for sleep across our natural circadian cycle (Borbély & Achermann, 1999) and the decreasing involvement of SWS across the sleep cycle as a whole (Walker & Stickgold, 2006).

Compelling evidence for this model was revealed in a study by Huber Ghilardi, Massimini, and Tononi (2004). Participants were trained on a motor adaption task requiring the movement of a handheld cursor to a target location relative to a covertly introduced clockwise rotational pressure. This task, compared to a no-rotation control condition, selectively activated regions in the right parietal lobe as measured by positron emission tomography (PET). Subsequent SWS activity not only increased locally at these regions, but also strongly correlated with waking improvements in task performance. Furthermore, SWS activity diminished over the first 90 minutes of NREM sleep, suggestive of a diminishing homeostatic pressure. These findings are in agreement with observations elsewhere noting increases in the post-training amplitude and slope of SWS activity (Sirota & Buzsáki, 2005; Sirota, Csicsvari, Buhl, & Buzsáki, 2003), and a reduction in slow oscillations when learning is prevented (Huber et al., 2006). Lastly, there is some evidence that sleep selectively improves performance on tasks once they are trained to high levels in both the procedural, (Hauptmann, Reinhart, Brandt, & Karni, 2005) and declarative domains (Tucker & Fishbein, 2008), in line with the predicted improvement in the SNR for particularly robust memory traces.

However, the synaptic homeostasis model by itself does not provide a complete explanation of sleep dependent consolidation. The specific mechanisms proposed to promote downscaling, by the authors own admission, remain somewhat vague (Tononi & Cirelli, 2003) and the inferred synaptic LTD during SWS remains somewhat at odds with the previously discussed literature implicating the same phase in LTP (King et al., 1999; Rosanova & Ulrich, 2005).

Despite this, it has been suggested that active consolidation and synaptic homeostasis are not mutually exclusive mechanisms, as outlined in a combined model dubbed *information overlap to abstract* (iOtA) (Lewis & Durrant, 2011). Here, it is proposed that sequential reactivations during SWS promote the selective retention for newly acquired memories with overlapping representations, while information with minimal overlap is simultaneously demoted through synaptic downscaling. Although this model neatly accommodates sleep dependent benefits for memories that share features (Ellenbogen, Hu, Payne, Titone, & Walker, 2007; Payne et al., 2009; Tse et al., 2007), its homeostatic component does not sufficiently explain the extent of sleep's involvement in processing weakly encoded memories.

1.3.5 The issue of memory strength and consolidation

Memories acquired under a range of impoverished, idiosyncratic or rare circumstances may still contain information that is paramount to survival (Norman, Newman, & Perotte, 2005) rendering an exclusively homeostatic account of consolidation potentially maladaptive. Although the issue of what exactly constitutes a 'strong' or 'weak' memory is itself a complex issue (Conte & Ficca, 2012), the literature has nevertheless consistently uncovered consolidation-associated benefits for memories retained at sub-optimal levels. Manipulations relevant to this notion include task complexity (Kuriyama, Stickgold, & Walker, 2004), memory interference effects (Ellenbogen et al., 2009), participant skill level (Peters, Smith, & Smith, 2006), and encoding depth (Drosopoulos, Windau, Wagner, & Born, 2007).

For instance in the procedural domain Kuriyama et al. (2004) manipulated the complexity of a finger tapping sequence to create a simpler (uni-modal) or harder (bi-modal) training procedure, with the latter version of this task conferring the greatest performance gains after sleep. Furthermore, the most difficult transitions between the sequence elements themselves, as indicated by diminished reaction times during training, displayed enhancements far in excess of those found for easier transitions.

Similar effects have also been found in declarative domain (Diekelmann, Born, & Wagner, 2010; Ellenbogen et al., 2009; Schmidt et al., 2006) including experiments conducted by Drosopoulos et al. (2007) that manipulated both the level of interference and encoding depth of word associations. Interference effects were generated by asking participants to learn two sets of arbitrarily associated words pairs, an A-B list (e.g. 'egg – pillow') followed by an A-C list (e.g. 'egg – shovel'). The order of list learning induced retroactive interference (RI) effects on the primary A-B list by virtue of the common 'A' item ('egg') shared with the secondary A-C list. A second manipulation trained participants to a weaker (≥60%), or stronger (≥90%), criterion during training. Sleep was found to not only facilitate the spontaneous recovery of A-B list information from RI, but also enhance the retention of weakly, but not strongly encoded, information. Using a similar paradigm it has also recently been established that these effects are not restricted to RI recovery, but may extend to the reduction of interference effects in general (Abel & Bäuml, 2013).

Although these findings might seem at odds with the previously noted research implicating the enhancement of highly robust memories (Hauptmann et al., 2005; Tucker & Fishbein, 2008) these observations are not necessarily anomalous. For example, for declarative memories the benefits of consolidation in this regard were only noted when training protocols were intermixed with testing, which might have led to a particularly strong priming of these materials for consolidation through mnemonic selectivity, as outlined in the previous section (Tucker & Fishbein, 2008). Moreover, if we are to consider combined models of memory consolidation, it may be that strong memories benefit more from synaptic downscaling, while weaker memories are more reliant on the reconstructive processes associated with systems-level consolidation, as outlined above (Diekelmann et al., 2009).

1.3.6 Criticisms of an active role for sleep in memory consolidation

Despite the evidence outlined so far the case for sleep-dependent consolidation has not remained uncontested (Siegel, 2001; Smith, 2001; Vertes, 2004). While several criticisms have already been addressed in this review two prominent issues will be explicitly outlined here: (i) methodological confounds inherent in sleep and memory research, and (ii) the existence of REM sleep consolidation.

Methodological issues in sleep research

Firstly, the derived methodology of some of the experimental research in this area may be prone to confounds relating to circadian or fatigue effects (Christianson & Christianson, 1992; Grosvenor & Lack, 1984). This relates, in part, to the lack of recovery sleep in some experimental conditions that manipulate the amount of sleep obtained by participants, as seen in the early/late paradigm used by Plihal and Born (1997), and circadian confounds inherent in the standard 12:12 sleep study design, which utilises a 9am to 9pm wake group and a 9pm to 9am sleep group (Tilley & Warren, 1983). In this latter case, participants may show faster reaction times yet lower accuracy levels in a given memory task undertaken in the afternoon or evening due to heightened metabolic activity associated with natural circadian pressures.

While these are legitimate issues the refinement of experimental methodology over the last 30 years has produced a range of increasingly sophisticated paradigms capable of dissociating the impact of sleep from circadian influences and the passage of time. This is well illustrated in a series of experiments conducted by Gais, Lucas, and Born (2006) who observed the benefits of sleep on vocabulary retention by supplementing a standard 12:12 design with various control conditions, including a night of sleep deprivation and measures of recall obtained 24 and 36 hours after uninterrupted periods of sleep or wakefulness. Other high profile investigations have similarly introduced robust controls to isolate the potential benefits of sleep on memory (Brashers-Krug et al., 1996; Huber et al., 2006; Karni et al., 1994; Wagner et al., 2004).

Additionally, the recent introduction of napping paradigms, where participants sleep for 60-90 minutes during the day or spend the equivalent time in resting wakefulness, have continued to generate positive results in favour of sleep-dependent consolidation while simultaneously minimising time-of-day effects (Mednick, Nakayama, & Stickgold, 2003).

Despite the above notions methodological issues are a continued cause for concern for sleep researchers. For example, clear behavioural readings can sometime be tricky to dissociate from *sleep inertia* which alludes to general cognitive deficits within the first 30-45 minutes after the onset of wakefulness. Moreover the ‘first night effect’ pertains to participants’ unfamiliarity with the sleep laboratory environment, a factor also known to detrimentally influence performance (Goerke et al., 2012). These issues in conjunction with the above caveats regarding 12:12 and 24 hour sleep designs necessitate thoughtful and replicable controls in a given experiment to address these, and other factors, which may potentially influence performance. Chapters 2-5 in the present thesis address such issues on a case-by-case basis.

The role of REM sleep in consolidation

Additional criticisms have been levelled at the role of the REM state in consolidation, as most prominently voiced by Vertes and colleagues (Vertes & Eastman, 2000a, 2000b; Vertes & Siegel, 2005; Vertes, 2004). These arguments primarily relate to the seemingly normal cognitive functioning of individuals with impaired REM sleep. Firstly, patients who survive damage to the pontine reticular formation, a brain region closely associated with the generation of REM sleep, will often display reductions in amount of time spent in this stage seemingly without any impairment to memory (Valldeoriola, Santamaria, Graus, & Tolosa, 1993). A similar observation has been made regarding individuals medicated on REM sleep suppressing antidepressants, such as selective serotonin reuptake inhibitors (SSRIs), who also do not seem to manifest any obvious deficits in memory (Thase, 1998). The issue of REM suppression therefore remains somewhat at odds with the position that REM sleep is essential for certain types of memory consolidation (Stickgold, 2005; Walker, Liston, Hobson, & Stickgold, 2002; Walker & Stickgold, 2010).

Counterarguments to these points relate to the relative lack of research systematically exploring the issue using validated consolidation-sensitive memory tasks (Ellenbogen, Payne, et al., 2006). Moreover, REM sleep is rarely suppressed entirely in these patient populations, raising the possibility that diminished REM activity is still sufficient for consolidation, or that other mechanisms within the sleep cycle can compensate. This latter eventuality was strongly implicated in a recent study exploring memory consolidation in rats dosed with the REM suppressing antidepressant desipramine (DMI) (Watts, Gritton, Sweigart, & Poe, 2012). Compared to a non-medicated control group these rats manifested an increased association between slow wave activity and memory performance, suggesting some form of neurological compensation to REM sleep suppression.

At present the anecdotal conjecture relating to the negligible impact of REM sleep is not enough to undermine the large body of experimental research repeatedly illustrating the importance of this stage in consolidation, particularly for emotional memories (Foulkes, Hollifield, Sullivan, Bradley, & Terry, 1990; Frankland, Bontempi, Talton, Kaczmarek, & Silva, 2004; Goerke et al., 2012; Ribeiro, Goyal, Mello, & Pavlides, 1999).

A final caveat should be made concerning studies that have found evidence for consolidation over wakefulness including the ongoing stabilisation (Brashers-Krug et al., 1996; Walker, Brakefield, Hobson, & Stickgold, 2003) enhancement (Atienza, Cantero, & Stickgold, 2004; Roth, Kishon-Rabin, Hildesheimer, & Karni, 2005) and integration of memories (Lindsay & Gaskell, 2012; Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010). It is by no means proposed that memory consolidation is a phenomenon exclusive to sleep and it should, rather, be considered a component of a larger set of dynamic processes encompassing both sleep and wakefulness (Dudai, 2004). Indeed, the previously discussed CLS model in section 1.2.2 makes it clear that waking repetition and reminiscence may also confer benefits in this regard (McClelland et al., 1995).

1.3.7 Summary and conclusions

The literature reviewed thus far provides a resolution to the active vs. passive debate introduced at the start of this section (Ellenbogen, Payne, et al., 2006) namely that these are not mutually exclusive hypotheses (Stickgold, 2005). While in some cases a merely passive-protective role for sleep cannot be ruled out, this does not in itself weaken the premise that sleep confers a unique biological state for memory consolidation. Indeed, while encoding and retrieval require wakefulness, memory processing may ultimately explain why we experience a loss of consciousness during sleep (Diekelmann et al., 2009).

Nevertheless, the issue must be raised regarding what exactly constitutes a ‘weak’ memory trace in terms of consolidation processes? This relates specifically to the evidence that system-level consolidation may help ‘salvage’ fragmented or weakly encoded memory traces. It is not presently clear from either an active consolidation or homeostatic perspective as to how we assimilate idiosyncratic information, particularly in relation to prior knowledge. For example, would information that is incongruent with what is already known be restructured, removed or incorporated more slowly into existing memory?

1.4 Sleep and memory transformation

While the majority of the literature discussed thus far deals predominantly with the quantitative products of consolidation, namely the enhancement and stabilisation of memory, recent research has begun to focus on the more qualitative changes memories can undergo as a result of the mnemonic reorganisation associated with consolidation. Such process may imbue memories with a range of properties that render them more applicable to future behaviour, thus underwriting a key adaptive benefit of consolidation (Stickgold & Walker, 2013; Walker & Stickgold, 2010). Dynamic consolidation in this manner is not as well documented as the more direct forms of mnemonic persistence discussed in the previous section, and are typically revealed through measures of integration that extend beyond veridical memory retention (Dudai, 2012; Inostroza & Born, 2013).

To this end, a number of innovative experiments have been developed to probe two such integrative dimensions: assimilation and abstraction. Assimilation refers to the integration of new memories into pre-existing knowledge structures (Walker & Stickgold, 2010), whereas abstraction denotes the extrapolation of invariant patterns of information from multiple overlapping memory traces (Lewis & Durrant, 2011). While these two modes of consolidation are closely associated, the former process relates specifically to the properties memories may be imbued with once incorporated into pre-existing memory, while the latter process relates to the formation of new concepts. Given that a primary benefit of semantic memory relates to its capacity to generalise information to new contexts and situations (Murphy, 2004; Tulving, 1972) both assimilation and abstraction may operate together to guide the uptake of relevant information over time (Inostroza & Born, 2013; Stickgold, 2013). Many of the dynamic changes to memory associated with these processes are conceptually outlined in Figure 6.

1.4.1 Memory assimilation

Assimilation, in terms of language acquisition, can be understood as two forms of lexical processing: ‘lexical configuration’ pertaining to the knowledge of factual information about words including phonology, orthography, and meaning, and ‘lexical engagement’, referring to a deeper level of processing, such as the integration of newly learned words into a pre-existing lexicon (Leach & Samuel, 2007). Lexical configuration, which can also be understood as the episodic knowledge associated with a given word, can manifest quickly, and endure up to a year after its initial acquisition (Salasoo, Shiffrin, & Feustel, 1985). To date our understanding of lexical assimilation in relation to memory consolidation is derived predominantly from measures of lexical engagement distinct from semantic integration.

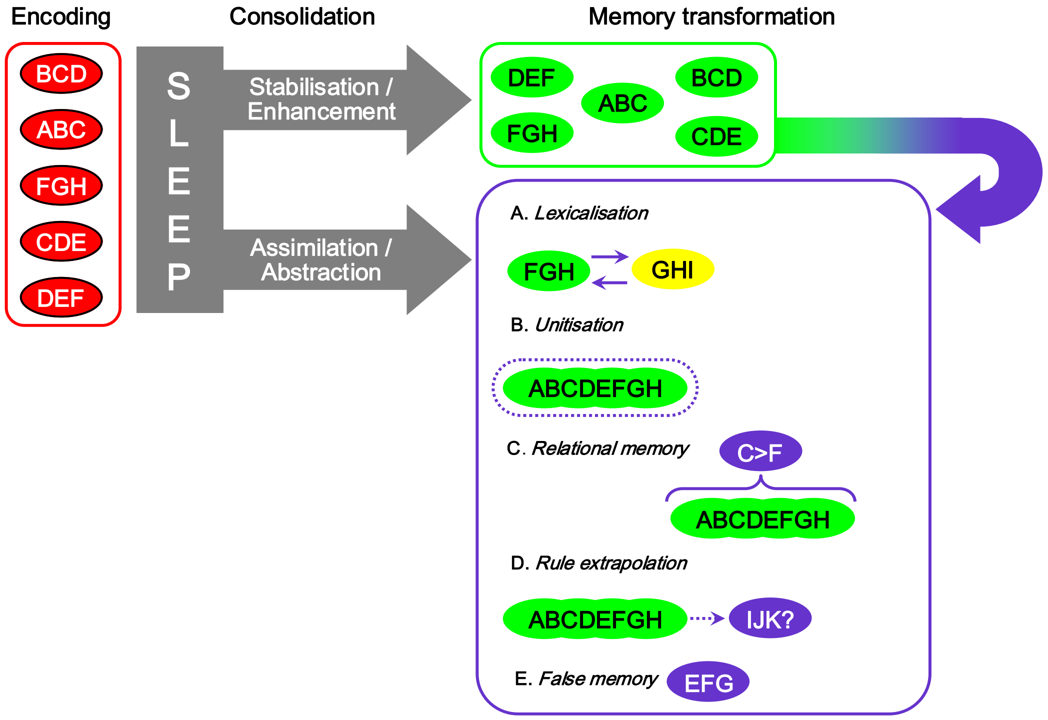


Figure 6. A figure conceptually detailing some of the different transformative changes memories can undergo through consolidation. The letter triplets in the red box represent potentially related pieces of new information at encoding. The products of consolidation are highlighted in the green and purple boxes. While the green box relates to the increased stabilisation and enhancement of veridical memory integrative transformations to the same memory traces are outlined in the purple box. These changes relate to (A) increased lexical competition with pre-existing vocabulary (the yellow oval), (B) the increased binding of overlapping memory elements into a single memory trace, (C) an enhanced capacity to infer relationships between recently acquired memories, (D) the extrapolation of novel rules from patterns of newly acquired information, and; (E) false memories that can arise from the extrapolation of a semantic theme across related pieces of information. This figure was partially adapted from Stickgold and Walker (2013).

These findings are well illustrated in a series of investigations examining the integrative effects, or ‘lexicalisation’, of newly learned novel word forms into existing vocabulary. Gaskell and Dumay (2003) generated a number of novel words items derived from real words already present in the English language, such as the novel word *cathedruke* from the associated real word *cathedral.* Over five days participants listened to these hybrid novel word items as part of a phoneme monitoring task, with the effects of immediate and prolonged exposure measured in terms of explicit (word recognition) and implicit (two-alternate forced choice) memory performance. Interestingly, although Gaskell and Dumay observed rapid explicit learning, more implicit integration only emerged after three days of training, whereupon recognition of the real word from which the novel word derived (e.g. *cathedral*) became slower in a lexical decision task. It was argued that integrative consolidation over time had induced a degree of lexical competition between novel words and their real word counterparts (Figure 5, A).

In a series of subsequent investigations the time course of such integration proved to be robust, lasting up to 8-months (Tamminen & Gaskell, 2008), while the period of offline consolidation required for these effects to become expressed was isolated to 24 hours (Dumay, Gaskell, & Feng, 2004) and eventually a single evening of sleep (Dumay & Gaskell, 2007). Subsequent fMRI analysis revealed that novel word acquisition initially resulted in strong hippocampal activation, with the strength of this activation predictive of lexicalisation (Dumay & Gaskell, 2007). Importantly, novel words eventually produced an activation profile in the neocortex similar to that of existing words, strongly suggestive of neurological assimilation in parallel with these behavioural changes. Additional studies have observed lexical competition after a period of sleep in both children (Henderson, Weighall, Brown, & Gaskell, 2012) and adults (Tamminen et al., 2010) with the latter investigation in particular revealing a relationship between sleep spindle activity and lexicalisation measures, implicating potential integration through system-level consolidation during sleep (Frankland & Bontempi, 2005; Rasch & Born, 2013).

1.4.2 Memory abstraction

The ability to abstract relationships across elements of episodic memory, and the capacity to interrelate this information to infer higher order relationships, are considered two additional integrative products of consolidation. The former process relates to the concept of ‘unitisation’, where it has been argued that overlapping memory representations in the hippocampus are consolidated and bound together to form a more unified memory trace (Figure 5, B)(Lewis & Durrant, 2011). As a consequence, or in parallel, with this process shared properties between these overlapping relationships may be more easily inferred (Walker & Stickgold, 2010) (Figure 5, C). Although this latter notion clearly relates to the strong relational properties inherent in semantic memory (Kumaran & McClelland, 2012) memory abstraction, as with lexical assimilation noted above, has primarily been observed for memories with little or no relevance to semantic memory.

This is perhaps best exemplified in a study by Ellenbogen, Hu, Payne, Titone, and Walker (2007) who deployed a transitive inference paradigm to probe consolidation in relation to flexible deductions across previously learned units of declarative information (Bryant & Trabasso, 1971). Participants were trained on a set of arbitrary stimuli that followed an implicit hierarchy (e.g. A>B>C>D>E>F). Training was implemented in a manner so that these items were only ever learned adjacently in ‘premise pairs’ (e.g. ‘A>B’, ‘B>C’... ‘E>F’). At test participants were asked to judge the order of novel combinations of stimuli requiring the extrapolation of information from the underlying premise pairs. These included inferences between items separated by one unit (‘B>D’ and ‘C>E’) and two units of information (‘B>E’) along the trained hierarchy. It was found that 12 hours of sleep or wakefulness produced equivalent improvements in performance on one unit transitive inferences, however judgements relating to the more distant two unit comparisons improved by 35% exclusively after a period of sleep. More recently an experiment replicated these effects, but only when training was reinforced, and not administered through passive exposure (Werchan & Gómez, 2013). This latter manipulation suggests transitive inferences in both of these studies may have derived from a form of procedural consolidation (Frank, Rudy, Levy, & O’Reilly, 2005; von Fersen, Wynne, Delius, & Staddon, 1991) rather than a qualitative shift in the expression of declarative memories as originally thought (Ellenbogen et al., 2007).

Nevertheless, mnemonic abstraction as a product of declarative memory consolidation has been demonstrated elsewhere. Prior to a sleep opportunity Lau, Tucker, and Fishbein (2010) presented participants with a series of face-object pairs, with two faces independently paired with the same object. After a 90 minute nap participants were required to judge which faces shared the same object, a task requiring the extrapolation of shared features across these items. It was found that participants who had slept not only manifested performance enhancements for both the original items and the novel comparisons, but also that the new relational judgments correlated with prior SWS (Lau et al., 2010). In a second experiment participants learned the English meaning of Chinese characters which shared overlapping semantic components called ‘radicals’. Although no correlations between sleep and memory were observed in this experiment, participants who slept were superior at judging new characters sharing the same radicals, as well as explicitly articulating the semantic properties shared across these items (Lau, Alger, & Fishbein, 2011).

While both studies further implicate sleep in the abstraction of information, the inconsistent involvement of sleep stage activity demands a closer examination of these effects in relation to more fine-grained components of sleep physiology, specifically sleep spindle activity (Stickgold, 2013). Furthermore, these investigations are not definitive, particularly as the key findings from the latter investigation were derived from ordinal scale applied to a seemingly nominal set of semantic errors (Lau et al., 2011).

Nevertheless, the above findings are in accordance with a small set of diverse studies elaborating the role of sleep in enhancing the abstraction of implicit rules and information (Figure 5, D). In terms of language learning infants have been found to preferentially orientate to novel, yet grammatically plausible, auditory strings derived from exposure to an artificial language prior to a nap (Gómez, Bootzin, & Nadel, 2006). Similarly, a period of sleep has been associated with an enhanced capacity to generalise perceptual elements of synthetic speech (Fenn et al., 2003) as well as the ability to generate novel associative links between unrelated word items (Cai, Mednick, Harrison, Kanady, & Mednick, 2009). Moreover, the abstraction of statistical frequencies in novel tone learning tasks suggests that participants are more adept at extrapolating novel rules after a period of consolidation, even if this information has no coherent semantic structure or meaning (Durrant, Cairney, & Lewis, 2012; Durrant, Taylor, Cairney, & Lewis, 2011). Lastly, a practical output of memory abstraction in this manner is perhaps evinced by increased levels of insight and awareness after a consolidation opportunity on a range of procedural (Stickgold & Walker, 2004; Wagner et al., 2004) and declarative tasks (Walker, Liston, et al., 2002).

1.4.3 Abstraction and integration in parallel

As noted at the start of this section the consolidation associated assimilation and abstraction most likely operate together to aid the continual refinement of predictive semantic memory (Stickgold & Walker, 2013). Indeed this notion forms a cornerstone of current theoretical accounts built around the purported bi-directional interaction between episodic and semantic memory systems during sleep (McClelland et al., 1995; McClelland & Rogers, 2003). Despite this theoretical impetus the impact of pre-existing knowledge (abstraction) on our ability to infer rules and patterns (assimilation) remains a largely untapped area in the field of sleep and memory research.

At present one paradigm that arguably lies at the intersection of these processes relates to the use of the Deese-Roediger-McDermott (DRM) false memory paradigm (Roediger & McDermott, 1995). Here, participants are taught a chain of semantically related words (e.g. ‘door’, ‘glass’, ‘pane’) with the most semantically salient word omitted (in this case ‘window’). A consolidation-associated enhancement in the rate of post-consolidation false memories for these critical word lures might be expected through the spreading activation of assimilated information during sleep (Anderson, 1983; Dell, 1986) or the extrapolation of thematic ‘gist’ across these associated words (Brainerd & Reyna, 2002; Reyna & Brainerd, 1995) (Figure 5, E).

Preliminary evidence suggests that sleep can indeed enhance false memories. Using the DRM paradigm Payne et al. (2009) found that, counter to the bulk of declarative memory research discussed so far, SWS was negatively associated with the recall of trained word items. Recall for false memory lures, on the other hand, increased by up to 50% relative to wakefulness. As false memory recall was enhanced relative to veridical recall it was proposed that consolidation in this manner might optimise memory for future use, with ‘gist’ superseding or even competing with the retention potentially superfluous episodic details. The findings of Payne et al., (2009) have been partially replicated in a study by Diekelmann et al. (2010) who found that false memory rate increased across periods of both sleep and overnight wakefulness, but only in low performing participants. This latter finding, and indeed the proposed generative mechanisms underlying false memories, are in line with the 'salvage function' of sleep noted in the previous section (Drosopoulos, Schulze, et al., 2007; Fenn et al., 2003). In this sense, a second benefit of assimilation and abstraction may be the elaboration of impoverished or incomplete memories into a broader semantic context.

However, despite these interesting findings the literature has not been completely consistent on this point. An experiment conducted by Fenn, et al. (2009) observed the opposite pattern of results, with nocturnal wakefulness and not sleep producing increased false memory rates. This was argued to be the result of diminished executive functioning in the sleep deprived pre-frontal cortex prioritising veridical enhancement over gist. Similarly, Diekelmann et al. (2008) found that sleep deprivation following learning, but not a period of sleep itself, increased false memory rates. To complicate the picture further, a recent fMRI study conducted by Darsaud et al. (2010) observed that sleep, relative to sleep deprivation, enhanced *both* veridical performance and false recollections, with no significant difference in the activation profile between these memory types.

So far the evidence relating to false memories has produced mixed, but not necessarily irreconcilable, findings. Firstly, there is good reason to believe that some of the differences between these investigations reflect variations in the methodology, rendering direct comparisons somewhat tentative (Straube, 2012). For example, Payne et al. (2009) measured false memories using recall, but not recognition tasks, with the former measure being potentially more sensitive to false memories as it relies on the internal generalization of cued features (Diekelmann et al., 2010; Inostroza & Born, 2013). Secondly, deprivation procedures may induce different motivational stressors and, potentially, the adoption of different strategies at retrieval. Although DRM memories have been argued to be resistant to varying participant strategies (Fenn et al., 2009) this does not appear to be the case after deprivation (Diekelmann et al., 2008).

As most of these studies did not monitor awareness levels or task insight several factors mediating false memory generation may have therefore been omitted. Nevertheless, as it stands, there is some evidence that sleep contributes to semantic integration. It is clear that this issue would benefit from a more standardised approach to false memory detection complimented by additional paradigms exploring this phenomenon.

1.5 Conclusions and thesis outline

While the evidence for sleep based consolidation suggests a rich and multifaceted role in the processing of declarative memories the sheer diversity of studies exploring the phenomenon brings with it a risk of fragmentation in the literature. As the sleep psychologist Robert Stickgold notes: “[…] one has to address the question of how five stages of sleep interact with at least six types of memories and six stages of post-encoding memory processing, for a combined total of 144 distinct sub-questions” (Stickgold, 2009, p. 305). The present thesis therefore brings together several of complementary approaches to consolidation outlined in this review to explore how the compatibility of new information in relation to prior-knowledge will influence its subsequent integration.

1.5.1 Research questions

Compatibility

The role of prior knowledge in consolidation will be explored by training materials in a manner that renders them congruent or incongruent with what is already known. Such a manipulation permits the delineation of potentially distinct integratory processes. For example, information that is congruent with pre-existing knowledge, according to the CLS and other models, should become rapidly consolidated, and should measurably benefit from the assimilative and abstractive qualities outlined in the previous sections (McClelland et al., 1995; Stickgold & Walker, 2004; Walker & Stickgold, 2010).

However, it is not clear from present theoretical and experimental accounts how consolidation operates on information that is incongruent with what is already known. It may be, for example, that such information will become restructured to conform to pre-existing schematic concepts through systems-level consolidation. This possibility fits well with notions in the present literature that consolidation aids in the reorganisation of new information to enhance its predictive value (Stickgold & Walker, 2013) as well as the recuperative benefits such processes have on more weakly encoded information, as noted in section 1.3.5. Alternatively, incompatible information may be more susceptible to rapid decay (i.e. forgetting) or actively suppressed through synaptic downscaling, active consolidation or both (Lewis & Durrant, 2011).

Mechanisms

The second issue the present thesis relates to the mechanisms underlying mnemonic compatibility and consolidation. A full exploration of the potentially divergent routes of consolidation outlined above necessitates a complementary exploration of sleep physiology.

Although the field of sleep and memory research has been highly productive, the delineation of sleep specific benefits in this regard has sometimes produced divergent results (Smith 2001). SWS has been associated with both the suppression (Payne et al., 2009) and enhancement (Plihal & Born, 1997) of semantically compatible information, while both REM sleep and SWS have been linked with the more integratory dimensions associated with memory consolidation (Lau et al., 2010; Stickgold, Scott, Rittenhouse, & Hobson, 1999). Similarly, underlying sleep spindle activity has been associated with multiple stages of consolidation including declarative enhancement (Cox et al., 2012; van der Helm et al., 2011) and mnemonic integration (Tamminen et al., 2010). Thus, while there is overwhelming evidence that sleep facilitates memory consolidation there is a need for experiments that attempt to distinguish the relative contributions of the sleep state in relation to some of these diverse observations.

It follows that information that is compatible with prior-knowledge should display the greatest relationship with SWS and its affiliated spindle activity, as these aspects of sleep physiology most directly relate to systems-level integration. However, as with the issue of incompatibility itself, it is not clear what mechanisms associated with sleep may be attributed to the processing of incompatible information. If this information is demoted through synaptic downscaling we may expect a negative relationship between the retention of this information and SWS. Alternatively, if systems-level dynamics aid in the restructuring and assimilation of incongruous information the resulting changes in memory should be positively associated with markers of systems-level integration during sleep. A third possibility is that the incongruent nature of such information will preclude any involvement of sleep-associated consolidation as it does not usefully contribute to predictive semantic memory. Here we might observe the rapid forgetting of incongruous information and no relationship with sleep physiology. These possibilities are more clearly outlined in the experimental chapters that follow.

1.5.2 Experimental chapters

To address these questions the present thesis contains five exploratory investigations across four experimental chapters. Broadly speaking these chapters will alternate between the two goals of this thesis: an exploration of (i) memory compatibility and consolidation, and (ii) the sleep-associated mechanisms involved with these processes. While Experiments 1 and 2 (chapters 2 and 3) focus on the role of abstraction, assimilation and consolidation, Experiments 3, 4 and 5 (chapters 4 and 5) probe the role of memory compatibility in the recovery of memories from interference. All of these experiments include multiple measures of integration in an attempt to capture some of dynamic changes that may occur to memory.

Experiment 1 implemented a modified transitive inference paradigm to explore the role of time on consolidation and semantic compatibility (Bryant & Trabasso, 1971; Ellenbogen et al., 2007). The experiment introduces a semantic dimension of size overlayed on premise pairs forming the underlying schematic hierarchy (e.g. 'A>B>C>D>E>F'). Real world items were interleaved with novel information during training which ran counter, or in line with, expected size relationships. Such a manipulation permitted useful comparisons between novel and pre-existing items in transitive judgments (e.g. 'B>D', and 'E>F'). Furthermore, the introduction of a semantic component to this task allowed generalisation measures to be formed in relation to additional real world items presented at test. While Experiment 1 explored consolidation in participants after 20 minutes (immediate) or 24 hours (delayed) Experiment 2 took elements of this design on to a 12:12 overnight sleep study monitored by PSG.

Experiments 3, 4 and 5 follow a parallel line of investigation by exploring the role of semantic compatibility in the recovery of memories from retroactive interference. Here, in a similar approach to Experiments 1 and 2, a 'classic' A-B, A-C retroactive interference paradigm was modified by introducing an expected relationship of size across lists (Koppenaal, 1963). Novel items ('A') were interleaved between two real word items ('B' or 'C') to create word triplets across lists that formed congruent or incongruent size relationships. This manipulation permitted an exploration of how the differential congruency of learned information may enhance or block the consolidation-associated recovery from retroactive interference (Drosopoulos, Schulze, et al., 2007). Experiments 3 and 4 explore this factor across 24 hours using immediate and delayed groups, with the night of sleep for delayed participants recorded using portable EEG devices. Experiment 5, as with Experiment 2, took this design on to an overnight study featuring PSG.

Chapter 2: Consolidation, prior knowledge and integration

2.1 Introduction

The ability to flexibly interrelate information is considered a core adaptation of the semantic memory system that permits generalisations and inferences to be formed about the world (Murphy, 2004; Rogers & McClelland, 2004). In Chapter 1 the standard model of memory consolidation was introduced which provides an established account of how predictive semantic networks might develop over time (McClelland & Goddard, 1996; McClelland et al., 1995; McClelland & Rogers, 2003). In this model contextually rich episodic information, stored primarily in the hippocampus, becomes decontextualised through its progressive integration into extra-hippocampal memory networks (Frankland & Bontempi, 2005; McClelland et al., 1995).

More recently the systems-level dynamics outlined in this model have been implicated in the ongoing assimilation and abstraction of new memories (Rasch & Born, 2013). The process of assimilation refers to the direct integration of new memories into pre-existing memory (Walker & Stickgold, 2010). Abstraction denotes a related process where information is extrapolated from recently acquired memories that share common or overlapping features (Lewis & Durrant, 2011). Despite the theoretical impetus highlighting the importance of both of these processes in the ongoing development and refinement of semantic memory (McClelland & Rogers, 2003; Rogers & McClelland, 2004) this possibility has yet to be comprehensively examined in relation to consolidation (Inostroza & Born, 2013).

To this end, the experiment presented in this chapter aims to explore the relationship between consolidation, memory compatibility and integration. This will be done by modifying a transitive inference paradigm, a task that provides a reliable measure of memory abstraction, to feature new information that is either compatible (congruent) or incompatible (incongruent) with pre-existing semantic knowledge. Moreover, the introduction of a semantic component to training will permit additional measures of assimilation (generalisation) not normally associated with this paradigm.

2.1.1 Transitive inference and consolidation

The capacity to establish indirect associations across linked episodic memories relates directly to our ability to perform *transitive inferences* (Eichenbaum, 2004; Lau et al., 2011, 2010). A transitive inference (TI) is said to occur when '...a reasoner co-ordinates two or more pieces of information to deduce a new conclusion, and those pieces of information overlap in a linear way.' (Wright & Howells, 2008, p. 244). For example, if you are independently told the relative heights of three people (Tom, Paul and Sarah) by first learning that Tom is taller than Paul (Tom > Paul), and later that Paul is taller than Sarah

(Paul > Sarah), a successful TI would be the conclusion that Tom is taller than Sarah (Tom>Sarah) despite having never directly compared the heights of these two individuals (Piaget, 1928, p. 87). Inferences in this manner are also thought to be an expression of *relational memory*, the capacity to flexibly interrelate stored elements of information (Konkel & Cohen, 2009).

Experimentally, the involvement of consolidation in the development of inferential reasoning has been established using the TI paradigm (Bryant & Trabasso, 1971; Dusek & Eichenbaum, 1997). A prominent 'six element' variant of this procedure, displayed in Figure 7, involves training participants on an implicit hierarchy of ordered, often arbitrary units of information ('A>B>C>D>E>F', where '>' means 'to be selected over') that are presented independently in premise pairs (e.g. 'A>B', 'B>C'...'E>F') that together form an overlapping set of hierarchical premises. At test, participants are asked to judge between novel combinations of this information separated by one ('B>D' and 'C>E') and two units ('B>E') across the schematic hierarchy. As these comparisons involve comparing elements of trained information in a novel manner, they are thought to provide one measure of inferential reasoning.

Interestingly, a consolidation opportunity has been found to have an beneficial effect on TI performance. Using the six element TI paradigm Ellenbogen, Hu, Payne, Titone, and Walker (2007) trained participants on a set of ranked textured patterns and then tested participants inferential capabilities either immediately after training, after 12 hours of sleep or wakefulness, or after a 24 hour delay. It was found that while a period of 12 hours of wakefulness or sleep was required for performance improvements to emerge for one unit TIs ('B>D' and 'C>E') only a time delay involving sleep conferred an exclusive 35% performance advantage for the more distant 'B>E' TIs. While these findings implicate a consolidation-associated enhancement of relational memory, the precise mechanisms underlying these behavioural changes remain unclear.

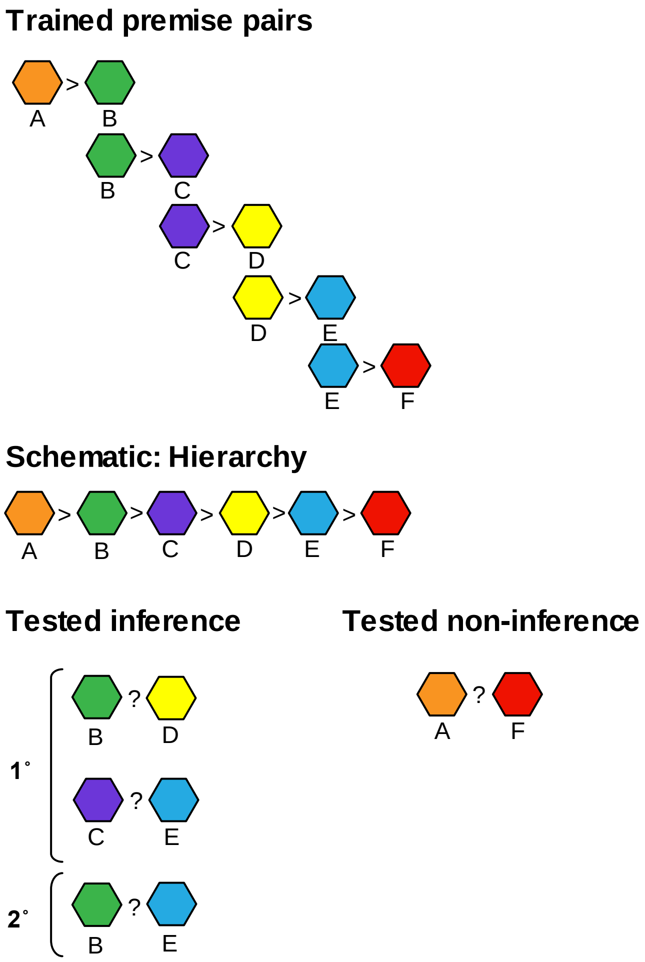


Figure 7. The transitive inference paradigm with six elements. A set of overlapping training pairs form an ordered hierarchy of relationships. Participants learn the individual premise pairs and are then asked to infer the correct order of novel combinations of this trained information. Items in inferential probe trials may be separated by one unit in the hierarchy (e.g., B ? D, indicated as 1°) or two units (e.g., B ? E, indicated as 2°). Non-inferential comparisons are also included to test knowledge of the relationship between the end items of the hierarchy (A ? F). This figure is adapted from Zeithamova et al. (2012).

On one hand, the observations made by Ellenbogen et al. (2007) implicate memory abstraction through systems-level consolidation. The hippocampal component of the standard consolidation model appears to support the rapid binding of conjunctive elements of information and the subsequent ability to flexibly co-ordinate inferences between these stored representations (Heckers, Zalesak, Weiss, Ditman, & Titone, 2004; Zeithamova, Schlichting, & Preston, 2012). However, recent computational modelling of the hippocampal system in relation to TI performance suggests that, although the hippocampus can indeed support the rapid enhancement of one unit TIs, the system in isolation cannot account for the slower emergence of two unit capabilities over time (Kumaran & McClelland, 2012). Instead, it has been proposed that the systems-level abstraction of recently acquired memories, and their subsequent assimilation into the neocortex, may enhance their relational properties after sleep (Lewis & Durrant, 2011).

That said, other literature relating to the TI paradigm suggests an alternative explanation. Specifically, the development of distal ‘B>E’ inferential capabilities is argued to be the product of reinforcement history rather than a synthesis of declarative memory *per se* (Frank, O’Reilly, & Curran, 2006; Frank et al., 2005). For example a *value transfer account* of TI proposes that B>E inferences develop from an inward transfer of reinforcement strength from the salient end items (‘A’ and ‘F’) to the affiliated ‘B’ and ‘E’ premises (von Fersen et al., 1991). Generalisation in this manner is actually typified by the absence of declarative awareness (Frank et al., 2005) and has been observed across a diverse range of non-human species including monkeys (McGonigle & Chalmers, 1977), fish (Grosenick, Clement, & Fernald, 2007) and pigeons (von Fersen et al., 1991).

Consequently, the findings of Ellenbogen et al. (2007) can be accounted for by the low-level strengthening of associations through synaptic memory consolidation, a process that does not necessarily involve the shifts in retrieval dependency associated with systems-level integration (Dudai, 2004; Hebb, 1949). This possibility was strongly suggested in a recent TI experiment conducted by Werchan and Gómez (2013) who trained participants on a series premise pairs using reinforced or non-reinforced training procedures. After a 12 hour interval involving either a period of sleep or wakefulness, it was found that only participants serving in the reinforcement condition displayed an improved ability to form ‘B>E’ transitive inferences. Moreover, as might be predicted by a value transfer account, consolidation in this manner occurred without any declarative awareness or insight (Werchan & Gómez, 2013).

Taken together, investigations exploring the development of inferential capabilities using the TI paradigm have, by definition, uncovered a degree of consolidation-associated memory abstraction. However, the role of declarative memory assimilation in these findings remains notably more contentious. Indeed, to date, the arbitrary nature of the stimuli used in previous TI investigations exploring memory consolidation (textured patterns) preclude any relevance to pre-existing semantic memory, thus potentially neglecting an important factor that may contribute to systems-level integration (Ellenbogen et al., 2007; Werchan & Gómez, 2013). One way to more thoroughly explore this issue is to introduce a semantic component to TI training.

As previously noted, such a modification to the TI paradigm will allow for a more comprehensive measure of memory assimilation, as the generalisation of newly learned materials can be probed in relation to pre-existing memory. Moreover, the issue of semantic compatibility itself can be explored, a factor considered relevant to mnemonic integration in current theoretical (McClelland et al., 1995; McClelland & Rogers, 2003; O’Reilly & Norman, 2002; Rogers & McClelland, 2004; Walker & Stickgold, 2010) and experimental accounts of consolidation (Diekelmann et al., 2010; McKeon, Pace-Schott, & Spencer, 2012; Payne et al., 2009; Tse et al., 2007, 2011).

2.1.2 Congruency and consolidation

The congruency of new information with prior-knowledge has become increasingly recognised as an influential factor in the selective integration of memory (Stickgold & Walker, 2013). For example, recent models have proposed that the top-down influence of the medial pre-frontal cortex (mPFC) during systems-level consolidation may be particularly important for blocking or facilitating the subsequent uptake of new information (Kroes & Fernández, 2012; van Kesteren et al., 2012).

This contention is supported by a recent fMRI study which manipulated the congruency (i.e. relevancy) of newly learned information by altering the coherence of expected plot-details in a short film clip (van Kesteren, Fernández, et al., 2010). Here participants were exposed to the first part of a short film in a scrambled (incongruent) or unscrambled (congruent) form prior to an offline delay. When the last 15 minutes of the film was presented to participants in an unscrambled form the following day, only individuals in the congruent condition displayed neurological markers of integration, as measured by a strong activation of the mPFC independent of the hippocampus. However, participants in the incongruent condition produced a strong co-activation between these two brain regions, suggesting that neurological integration had not taken place. Here, the mPFC was argued to induce a ‘fast-tracked’ route of systems-level integration for congruent memories.

These findings are in line with additional experimental work highlighting the rapid behavioural and neurological assimilation of congruent information. It has been shown that rats will rapidly generalise the memories of food-reward locations trained in a schematically consistent manner after a delay (Tse et al., 2007). Moreover, this generalisation was seen to occur in conjunction with a shift in retrieval dependency from the hippocampus to the neocortex. These markers of behavioural and neurological integration did not extend to inconsistent schematic training that occurred in parallel. Lastly, congruency with prior knowledge was manipulated in an fMRI investigation exploring the retention of words associated with object-fabric combinations after a consolidation opportunity (van Kesteren, Fernández, et al., 2010). When conjunctive elements of information were compatible with prior knowledge (e.g. a ‘leather jacket’) memory performance improved over 24 hours, and produced an activation profile suggestive of accelerated integration. Conversely, when object-fabric combinations contradicted expectations (e.g. ‘a lace umbrella’) minimal integration was found.

Although the present experiment is strictly behavioural, the above literature indicates that when new information is trained in a manner that is congruent with prior knowledge we should expect the relatively rapid assimilation of this information, particularly in relation to equivalently trained incongruent information.

2.1.3 Experiment aims and predictions

The present experiment aims to provide a more definitive exploration of how consolidation and prior knowledge may influence the integration of declarative memory. This was done by adapting the six element TI paradigm present in Figure 7 to probe memory abstraction and assimilation after a variable offline delay of 20 minutes (the immediate condition) or 24 hours (the delayed condition).

The congruency of information was manipulated by replacing the typically arbitrary premise pairs underlying the schematic hierarchy (e.g. 'A>B', 'B>C' etc.) with novel and real word paired associates. Here, the real word items provided explicit information regarding the relative size of the adjacent novel words, thus creating a unified size dimension across premises (e.g. 'giraffe>swock', 'swock>sheep' etc.). Interleaving novel words and real words in this manner will allow participants to infer the size of novel words from adjacent real word information. Consequently, the 'usefulness' of the overall size dimension was manipulated so that novel items were nested between real words that were either informative (i.e. *congruent* with prior knowledge) or contradictory (i.e. *incongruent* with prior knowledge).

Three measures of integration were established based on the behavioural properties of the TI paradigm in order to assess the impact of congruency on consolidation. While this particular experiment is not designed explore the sleep-specific mechanisms contributing to integration, all participants in the delayed condition will obtain sleep, making such accounts relevant to the following predictions. Moreover, in line with the research outlined in the previous section, an overarching hypothesis framing the present investigation is that memories derived from information that is congruent with prior knowledge will become rapidly integrated relative to incongruent information over 24 hours (Stickgold & Walker, 2013; Wilhelm et al., 2011). The manner in which this integration may be expressed is also predicted to differ across the three behavioural measures featured in this experiment.

Premise pairs and the serial position effect

The first integratory measure relates to memory of the premise pair items underlying the schematic hierarchy. Performance on this measure will be probed in relation to the *serial position effect,* a behavioural phenomenon induced by differences in the reinforcement history of premises during training (Greene, Spellman, Dusek, Eichenbaum,

& Levy, 2001; Moses, Villate, & Ryan, 2006; Ryan, Moses, & Villate, 2009; Wynne, 1995). Specifically, as the ‘A’ and ‘F’ items are always trained as correct and incorrect respectively, memory for the end premises (‘A>B’ and ‘E>F’) is typically stronger than central premises (‘B>C’, ‘C>D’ and ‘D>E’). The net result of these reinforcement differences at test is a serial position curve marked by weaker memory retention for central items. As outlined in Figure 8 below, the manipulation of congruency is predicted to have a differential impact on the expression of the serial position effect over time.

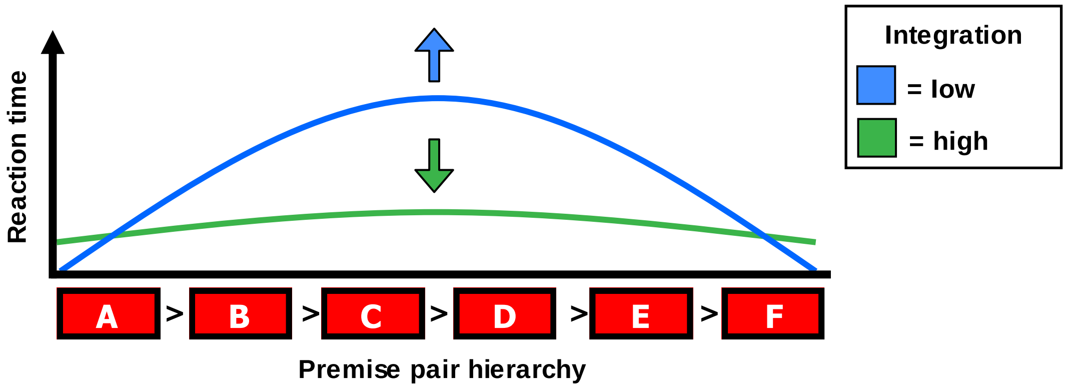


Figure 8. The serial position effect and memory integration. The serial position effect refers to the observation that memory performance for premise pair items in the transitive inference paradigm is typically stronger for items at the end of the schematic dimension (A>B and E>F). The extent of the serial position effect (the green and blue reaction time curves) is predicted to be influenced by the semantic congruency of the schematic hierarchy (see text).

For congruent information the serial position effect may be mitigated as premises are rapidly assimilated into pre-existing schematic knowledge (the green line in Figure 8). Firstly, should consolidation processes result in increased access to pre-existing semantic knowledge then the reinforcement history ‘bookending’ the schematic hierarchy may become less influential. This may occur as integration extends the otherwise finite size limits framing the trained hierarchy into a broader dimension of size. A related eventuality is the process of ‘unitisation’ outlined in the previous chapter. Combined models of consolidation have proposed that elements of episodic information that share overlapping features will become increasingly bound together into a singular memory trace through the cumulative reactivation of these shared components during sleep (Lewis & Durrant, 2011). As the congruent premises in the present experiment derive from both an overlapping semantic and associative dimension we would predict particularly strong unitisation effects after a delay. Together, or independently, these processes may ‘equalise’ performance across the trained premises, thus flattening the serial position curve.

By extension the influences of consolidation noted above are not predicted to promote the integration of incongruent information. Moreover, as the incongruent semantic training will result in less overlapping features across premises any resulting memory traces may be more vulnerable to disruption, decay or suppression over time (Wixted, 2004). If this eventuality is the case we may expect relatively poor memory performance on these items in the immediate condition which might become progressively worse after a 24 hour delay, resulting in the preservation, or even exacerbation, of the serial position effect over time (the blue line in Figure 8).

Transitive inferences

As the central premises in the present experiment contain interleaved sequences of real and novel words (e.g. ‘swock>sheep>creal>lizard’) useful comparisons can be made between one unit transitive inferences containing only novel words (e.g. ‘swock>creal’) and real words (e.g. ‘sheep>lizard’). Real word TIs will provide a useful baseline measure of inferential performance between word items already established in semantic memory. If congruent novel words are rapidly assimilated into pre-existing memory then we would predict an enhancement in TI performance for these comparisons after 24 hours, particularly in relation to the latter baseline measure.

Size generalisations and the semantic distance effect

Lastly, the introduction of a size dimension across the trained premises provides the opportunity to explore how novel words may generalise to untrained yet familiar word items. This measure relates to the emergence of the *semantic distance effect*, the long-standing observation that size judgements between items closer in size (e.g. ‘rat > frog’) will produce slower responses than more distant comparisons (e.g. ‘rhino > frog’) (Dean, Dewhurst, Morris, & Whittaker, 2005; Moyer, 1973). This effect is thought to derive from mental comparisons between items that exist along an internal psycho-physical dimension of size. Comparisons between items that are closer together on this dimension are thought to be less visually distinct, and accordingly harder to distinguish, rendering proximal size comparisons slower than more distal comparisons (Pavio, 1975).

To the author’s knowledge the semantic distance effect has not been explored in relation to the integration of novel words, or indeed inferred (as opposed to pre-existing) sizes distinctions. Nevertheless, a speculatory hypothesis illustrated in Figure 9 is that the assimilation of congruently trained novel words may lead to a stronger expression of the semantic distance effect over time. This may be the product of heightened imagability given the that systems-level consolidation is thought to promote the integration of new information into pre-existing semantic memory (McClelland & Rogers, 2003; Rogers & McClelland, 2004). If this is indeed the case, then comparisons between pre-existing words and novel words of a similar size should become slower in accordance with the semantic distance effect. Conversely, more distant size comparisons may become increasingly faster over 24 hours (the green line in Figure 9). Given the lack of predicted assimilation of novel words in the incongruent condition, the size distance performance curve associated with these responses should remain relatively flat on this measure (the blue line in Figure 9).

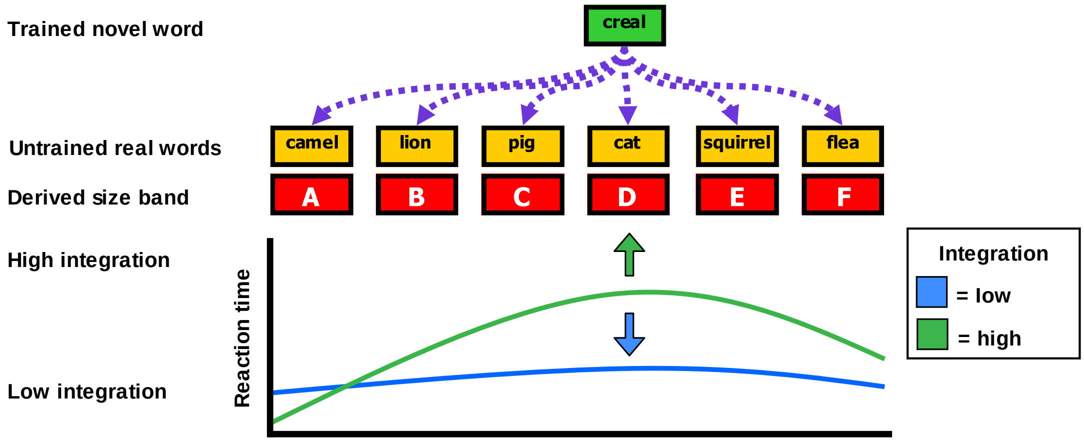


Figure 9. The semantic distance effect and memory integration. Size judgments between items closer in size are typically faster than judgments between items more distant in size (Moyer, 1973, Pavio, 1975). Novel words more integrated into a pre-existing semantic size dimension should display slower reaction times for similarly sized items compared to more distant items (high integration). Conversely, if little or no integration has taken place we might expect a flatter, and slower, reaction time curve across generalisation comparisons (low integration). Coloured arrows indicate the direction of these effects in terms of integration.

2.2 Methods

As the methods used in the experiment are very similar to those featured in Experiment 2 (Chapter 3) they will be outlined once in detail here with only deviations from these procedures presented and justified in the following chapter.

2.2.1 Participants

Forty-two participants aged between 18 and 29 were recruited through the University of York Psychology electronic experiment booking system (PEEBS). All were healthy, non-smoking native English speakers with no prior history of sleep or psychiatric disorders. All participants agreed to abstain from caffeine and alcohol over the course of the experiment and 24 hours beforehand. Written informed consent for participation was obtained in compliance with the local ethical committee.

Seven participants were removed from analysis for failing to complete training (n=5), failure to follow experimental protocol (n=1), and for producing excessively slow RTs at test (>2.5 standard deviations from the group mean, n=1). The remaining 35 participants were randomly assigned to Immediate (n=19; eight females) or Delayed (n=16; nine females) testing sessions. After randomisation the mean age of the Immediate condition was 20.3 years (*SD*=1.4) and 21.3 (*SD*=2.8). Overall, participants had a mean age of 21 years (*SD*=2.17) and were provided payment or course credit for taking part, with an additional £20 Amazon gift voucher awarded to the best overall performer.

2.2.2 Design

The present experiment followed a mixed-design featuring the between-participant variable of delay type, with groups tested 20 minutes (immediate) or 24 hours (delayed) after training (Figure 10, A). The within-participant variables relate to Congruency (Congruent or Incongruent) and measures of integration relating to the serial position effect, TIs, and the semantic distance effect (as defined below).

The order of training and testing tasks can be found in Figure 10, B. The training session consisted of a novel word familiarisation phase followed by a premise pair size judgement phase. The testing session consisted of two separate size judgement phases: (i) re-exposure to the original trained premises intermixed with novel TI comparisons, and; (ii) a generalisation phase requiring comparisons between novel items against a range of familiar, but untrained real word items. Training and testing sessions were staggered across the day (9am-5pm) to minimise the potential influence of time of day effects (Tilley & Warren, 1983).

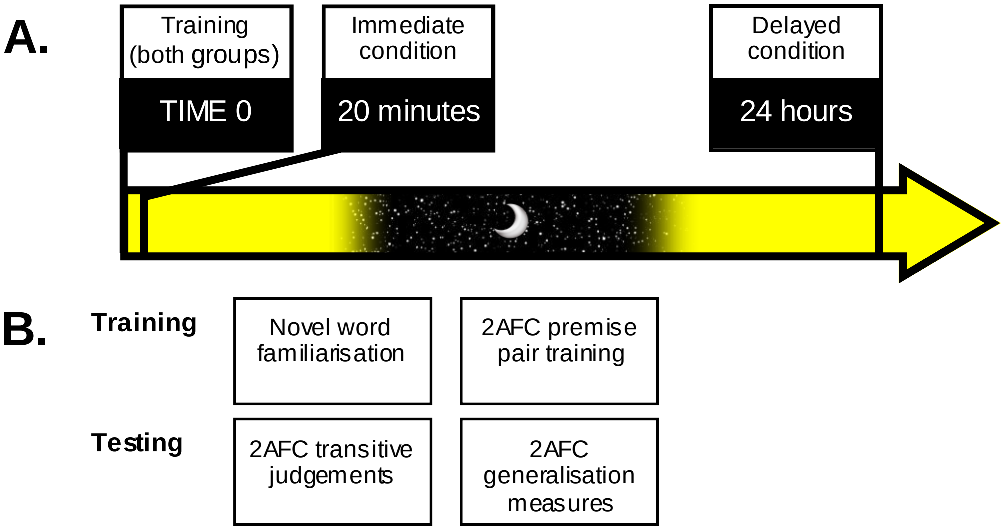


Figure 10. Experiment 1 design. (A) The between participants conditions in this experiment. Testing was conducted 20 minutes (immediate) or 24 hours (delayed) after training. (B) The sequential order of training and testing tasks. 2AFC refers to the two alternative choice size judgements (see procedure).

2.2.3 Stimuli

Overview

Stimuli consisted of two sets of syntactically plausible novel word items interleaved with real word items relating to a theme of land dwelling animals or household objects. These themes were overlaid onto a six element schematic hierarchy (e.g. ‘A>B>C>D>E>F’) following an ordinal dimension of size. For example, in their congruent form, the animal theme comprised of ‘giraffe > swock > sheep > creal > lizard > floth’ while the congruent object theme comprised of ‘fridge > hetel > shirt > gadge > book > tevel’ (Figure 11).

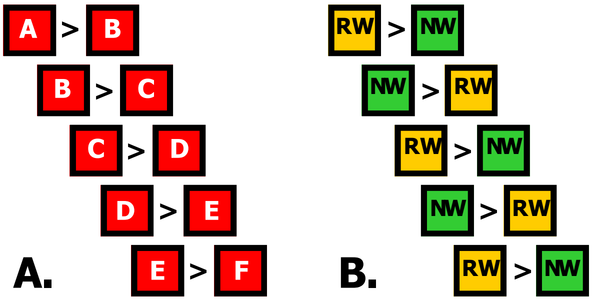


Figure 11. Experiment 1: The adapted transitive inference paradigm. (A) An example of a typical six element premise pair hierarchy. (B) The premises in the present experiment adapted to include pairs of novel words (NW) interleaved with real words (RW) that convey a semantic dimension of size.

Congruency

Congruency was manipulated by shifting the location of real word items so that the semantic dimension was either congruent or incongruent with the associative reinforcement across the schematic hierarchy (Figure 12). Note that this manipulation altered the real word, but not novel word, placement in the schematic hierarchy.

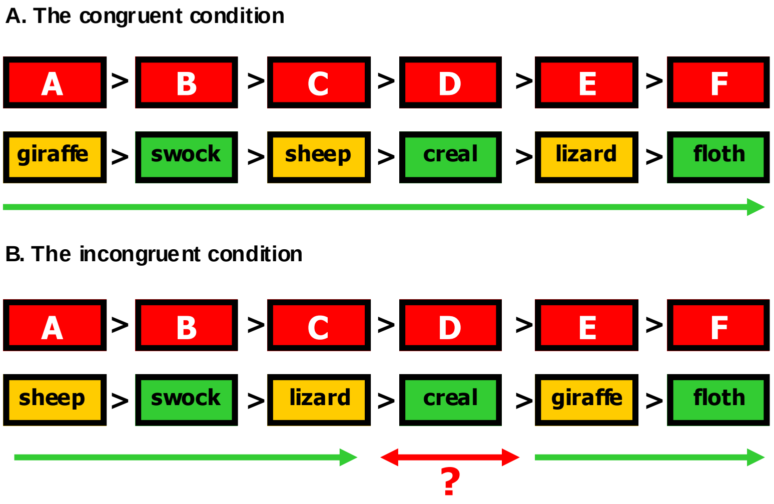


Figure 12. Experiment 1: The congruency manipulation. Red boxes indicates the underlying premise reinforcement which remained consistent across congruency conditions. Yellow boxes indicate real word items while green boxes represent novel words. (A) In the congruent condition the semantic dimension (the green line) was in accordance with the underlying premise reinforcement. (B) In the incongruent condition this semantic dimension was disrupted by altering the positions of real words within the hierarchy. Specifically, the novel item ‘D’ received conflicting real word size information (the red line). Note that for illustration purposes this example displays the animal theme only and participants were in fact trained on two independent hierarchies counterbalanced in terms of theme (animals or objects).

Transitive hierarchy construction

Real words

Real word items were rated on their size and typicality by 20 native English speaking individuals who did not take part in the main experiment. Viable animal (n=35) and household items (n=27) were placed in a ranked order of rated size and divided into six groups to form size bands ranging from the largest (e.g. ‘elephant’ and ‘car’) to smallest items (e.g. ‘flea’ or ‘lightbulb’). Words central to three of these size bands were allocated to the equivalent ‘A’, ‘C’ and ‘E’ premises in the hierarchy while premises ‘B’, ‘D’ and ‘F’ were left open as ‘slots’ for novel words (see below). Please refer to Appendices 1 and 2 for specific details regarding this hierarchy construction.

Novel words

Novel words were generated using the software program Wordgen (Duyck, Desmet, Verbeke, & Brysbaert, 2004) and were comparable to the real word items on several factors relevant to memorability. This included equivalent measures of bigram frequency between real words (*M*=7262, *SD*=3332) and novel words (*M*=7286, *SD*=3414) and the matching of the average syllable count between real word (*M*=1.4, *SD*=1.7) and novel word stimuli (*M*=1.4, *SD*=0.4). Additionally, as the average word length of the real word items was 5.3 letters (*SD*=1.3), all novel items were five letters long.

Viable novel words were randomly allocated to the ‘B’, ‘D’ and ‘F’ premises in the schematic hierarchy thus falling between real word items in terms of size (e.g. ‘giraffe > swock > sheep’)(Figure 13). Two non-hierarchical filler pairs (Figure 13, to the right of the purple dashed line) were included in training to avoid overtly revealing the underlying size continuity, a factor that might otherwise lead to rapid ceiling performance (Frank, Rudy, Levy, & O'Reilly, 2005; Libben & Titone, 2008). Real word filler items were derived from the rated ‘B’ and ‘D’ size bands (Appendix 1).

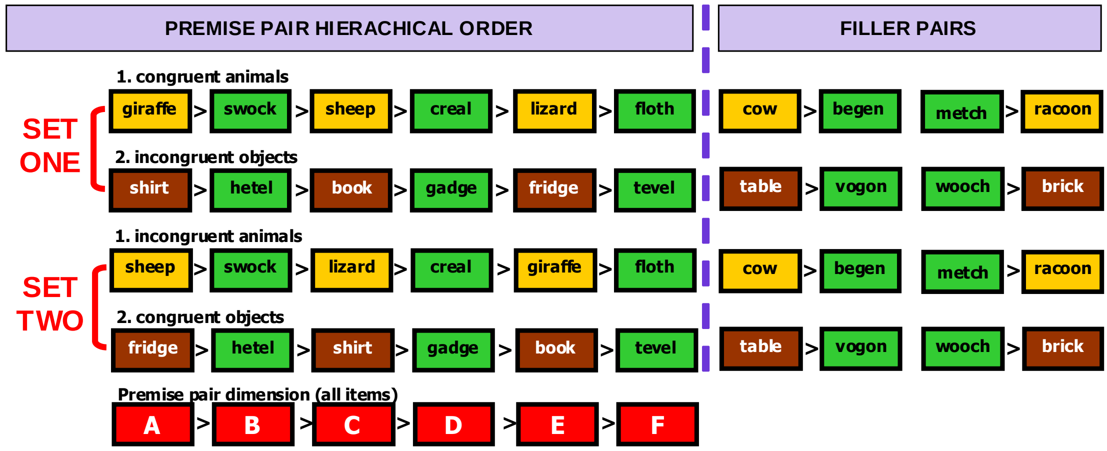


Figure 13. Experiment 1: The counterbalanced congruency conditions featured in training. Each participant learned one set (red brackets) of congruent or incongruent animal and object hierarchies. The order of presentation (animals or objects trained first) was also counterbalanced for a total of four potential training sets. The filler pairs to the right of the purple dashed line contain isolated yet thematically related size judgements included to mask the primary schematic dimensions.

2.2.4 Procedure

All instructions and experimental tasks were presented in a quiet testing room on a 17.1 inch computer monitor using E-prime software (Psychology Software tools, Inc., Pittsburgh, PA).

Instructions

A scenario for training was introduced to provide a context for the stimuli and format of the experiment (see Kumaran, Summerfield, Hassabis, & Maguire , 2009). Participants were informed that they were part of a science team in a newly discovered area of jungle and that, while they remained at base camp, field scientists would relay information back to them regarding newly discovered animals and objects named in a local language. For classification purposes these items would be judged in terms of their size relationships with familiar objects and animals but, due to an equipment failure, this information would only be relayed back to the participants in pairs of words requiring trial and error learning. The complete set of instructions for this scenario can be found in Appendix 3.

Training procedure

Novel word familiarisation

To ensure prior attendance to the novel word stimuli participants underwent a pre-training familiarisation phase at time 0 (Figure 10). Ten novel word items were presented four times in a randomised order as follows: (i) a 600 msec blank screen, (ii) a 600 msec fixation point, (iii) a 600 msec novel word presentation, followed by; (iv) an untimed text field with instructions to type the previously presented and press the ‘return’ key when finished. This cycle repeated until 40 trials had been completed (~5 minutes total).

Premise pair training

Participants were pseudorandomly assigned to one of four potential training sets (as outlined in Figure 13) to ensure an equal distribution between all congruency and training order combinations. Participants were exposed to seven premise pairs (five transitive, two filler) per training block with the order of on-screen presentation (left or right) counterbalanced to avoid a location response bias. Accuracy and RT scores were recorded using the left and right buttons on a Cedrus model RB-520 button box with participants choosing the side of the screen corresponding to the animal or object they perceived as the largest.

Trials within each block were presented in a pseudorandom manner to avoid overtly revealing the schematic hierarchies (e.g. ‘sheep>swock’ would not precede ‘swock>lizard’ to avoid making the order ‘sheep>swock>lizard’ obvious). Each trial was presented as follows: (i) a 600 msec blank screen, (ii) a 600 msec fixation point, (iii) an untimed 2AFC size judgement within a premise pair, and; (iv) 1000 msec of feedback. This cycle was repeated for a total of 14 trails per training block (~1.5 minutes).

As with other studies of this type, associative reinforcement for correct and incorrect responses was provided after each trial (Acuna, Sanes, & Donoghue, 2002; Ellenbogen et al., 2007; Werchan & Gómez, 2013) accompanied by overall percentage accuracy scores at the end of each block. Furthermore, a hidden ≥80% performance threshold was activated after the third training block to exit training. This exit protocol is a standard implementation in the literature intended to compensate for individual differences in learning efficiency (Ellenbogen et al., 2007; Frank et al., 2005; Siemann & Delius, 1996). Lastly, animal and object hierarchies were trained independently based on pilot data indicating that when thematic blocks were interleaved (e.g. ‘animal, object, animal’) participants became confused and struggled to successfully complete training.

Once this session concluded (~20 minutes total) participants either left the testing environment to return after 24 hours (the delayed condition) or stayed and surfed the internet for 20 minutes (the immediate condition). To minimise the chance of rehearsal between training and testing participants were mildly deceived regarding the subsequent testing phase, being told initially that there would be a second round of training.

Testing procedure

Transitive inference testing

The format for the surprise testing phase was the same as premise pair training except that all reinforcement cues and exit criterion were removed and the five trained premise pairs were reintroduced alongside four new comparison pairs. These new trials included two inferential pairs separated by one unit of information between (i) two novel words (e.g. ‘B?D’ or ‘swock?creal’), and; (ii) two real words (e.g. ‘C?E’ or ‘sheep?lizard’). A third more distant inferential pair was formed between the items separated by two units of information (e.g. ‘B?E’ or ‘swock?lizard’). Lastly, items at the extreme ‘end anchors’ of the hierarchy (e.g. ‘A?F’ or ‘giraffe?floth’) acted a non-inference control comparison. Knowledge of the central hierarchy is not thought to be required to solve these latter comparisons due to the strong reinforcement history of the ‘A’ and ‘F’ items, typically resulting in near ceiling levels of performance (Wynne, 1995).

Participants at the start of the test phase were informed that there might be novel comparisons and they were to make their “best guess” regarding the correct answer. TI testing occurred across four interleaved blocks of object and animal items. Each block contained nine comparisons (five premise pairs, three TIs and one non-inference comparison) presented twice in a counterbalanced fashion (e.g. ‘swock?creal’ and ‘creal?swock’) for a total of 18 trials per block presented as follows: (i) a 600 msec blank screen, (ii) a 600 msec fixation point, followed by; (iii) an untimed 2AFC size judgment (~15 minutes total).

Semantic distance testing

The final testing phase required participants to make 2AFC size judgments between all of the trained novel words and new, untrained real word items derived from all six size bands (A, B, C, D, E and F; Appendix 1). As such each novel word (e.g. ‘creal’) was compared to a range of sizes from the same theme (e.g. ‘camel’, ‘lion’, ‘pig’, ‘cat’, ‘squirrel’ and ‘flea’). Furthermore, on screen counterbalancing (‘A?B’ and ‘B?A’) used two real word items derived from the same size band (e.g. ‘lion?creal’ and ‘creal?bear’), as opposed to the same item twice, as the first presentation might otherwise prime the second. Each trial was presented as follows: (i) a 600 msec blank screen, (ii) a 600 msec fixation point, and; (iii) a 10000 msec timed 2AFC judgement. Each novel word (6 in total) was therefore compared to two animals from each size band (n=24) for a total of 144 comparisons spread equally across two blocks (~15 minutes total). Although items from different semantic categories were never directly compared, animal and object judgements were intermixed in this testing phase to increase demands on semantic flexibility and integration.

Other measures

A sleep log (Appendix 4) was introduced to probe the amount and quality of sleep, and to screen participants. Sleep quality was rated on a 1-10 scale (worst to best). This questionnaire was completed before the first testing phase for all participants.

The Stanford Sleepiness Scale (SSS) was introduced before both training and testing phases to monitor comparative levels of self-rated alertness. This itemised scale ranges from 1 (highly alert) to 7 (sleep onset) (Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973)

(Appendix 5). The decision to use these measures was based on advice from Dr. Jakke Tamminen who suggested that the validity of the SSS would be improved with the implementation of a supplementary visual analogue scale.

Lastly, participants were screened for potential confounding issues using a series of prepared questions in a phone call including the requirements: (i) no history of mental health problems, (ii) no use of any psychiatric or mood altering medication, and (iii) English as a first language. These were later followed up by a paper questionnaire asking the same questions prior to the experiment.

2.3 Results

All data were analysed using SPSS version 18.0 with a two-tailed $p<.05$ threshold for significance. All RT results are reported in milliseconds (msec). Extreme fast (<300 msec) and excessively slow (>5000 msec) RTs were removed from the data prior to analysis (3.3%).

2.3.1 Group equivalency

The two delay types were compared on a number of measures to ensure equivalency in terms of training performance, sleep patterns and levels of alertness.

Training

The number of trials to criterion during training (between 4 to 10) and percentage performance on the final block across all items (≥80%) were each analysed using a 2x2 mixed-effects ANOVA featuring the between-participants variable of Delay type (Immediate vs. Delayed) and the within-participants variable of Congruency (Congruent vs. Incongruent).

The products of this analysis, which can be found in Table 1, revealed no main effects or interactions for any of these factors regarding the number of training blocks required to exit training (M=5.2, SE=.22), Fs<2.67, all ps>.112, or accuracy on the final exit block (M=92%, SE=.7), Fs<1.88, ps>.179. Analysis was repeated for exit accuracy performance restricted to the three central premise pairs (B>C, C>D, D>E) as these comparisons may form the basis of subsequent transitive inferences (Bryant & Trabasso, 1971). Again, no main effects or interactions emerged (M=86%, SE=.17), Fs<1.38, ps>.249. As such, training performance remained undifferentiated by Delay type.

Table 1. Experiment 1: Training performance.

	Immediate condition			Delayed condition		
Congruency	Blocks until exit	Accuracy all premises (%)	Accuracy central premises (%)	Blocks until exit	Accuracy all premises (%)	Accuracy central premises (%)
Congruent	5.1 (.3)	91 (1.5)	83 (3.1)	4.9 (.4)	93 (1.6)	91 (3.4)
Incongruent	5.2 (.3)	92 (1.1)	84 (3.1)	5.8 (.4)	90 (1.2)	84 (3.3)

Note: Accuracy for all premises includes filler items. Central premise accuracy refers to C>D, D>E and E>F items only. Values in parentheses denote the standard error.

Sleep measures

Unpaired t-tests were conducted between Delay type (Immediate vs. Delayed) on two self-reported variables relating the sleep obtained prior to the testing phase. Specifically (i) the amount of sleep reported on the night prior to testing (in hours), and; (ii) the perceived quality of this night of sleep as rated on a scale of 1-10 (from worst to best). The amount of sleep prior to testing did not differ between Delay type (M=7.8, SE=.27), t<1, nor did the rated quality of sleep (M=7.5, SE=.34), t(33)=1.58, p=.123, indicating consistent sleep parameters between the Immediate and Delayed conditions.

Alertness levels

Lastly, to compare levels of alertness during the experiment SSS ratings were entered into a 2x2 mixed-effects ANOVA featuring the between-participant variable of Delay type (Immediate vs. Delayed) and within-participant variable of Phase (Training vs. Testing). A marginally significant trend of Phase emerged with all participants showing a tendency to rate themselves as less sleepy at training (M=2.48, SE=.1) than at testing (M=2.72, SE=.14), F(1,33)=3.7, p =.061, however this effect was not qualified by any additional main effects or interactions between the two factors, Fs<1.2, ps>.283, indicating that SSS ratings throughout the experiment were equivalent between the two time delays (M=2.6, SE=.11) (Hoddes et al., 1973).

2.3.2 Premises pair performance and the serial position effect

The first integratory measure relates to the memory of central premises in relation to the serial position effect. To measure the extent of this effect performance for end (A>B, E>F) and central (B>C, C>D, D>E) premise pairs were independently averaged together to form the factor of Premise type (End vs. Central). These measures were entered into a 2x2x2 mixed-effects ANOVA featuring the between-participant variable of Delay type (Immediate vs. Delayed) and within-participant variables of Congruency (Congruent vs. Incongruent) and Premise type (End vs. Central) using the dependent variables of accuracy (%) and RTs (msec). It should be noted that RTs, as opposed to accurate RTs, were used in this analysis due to the chance accuracy levels of some of the incongruent measures (see Table 2)

Accuracy levels and the serial position effect

The analysis of accuracy performance, which can be found in Table 2 revealed three main effects. Firstly, a main effect of Delay type was present, *F*(1,33)=13.62, *p*=.001, driven by higher accuracy levels in the Immediate condition (*M*=78%, *SE*=1.82) relative to the Delayed condition (*M*=68%, *SE*=1.99). Secondly, a main effect of Congruency was apparent, *F*(1,33)=6.12, *p*=.019, indicating higher overall accuracy levels for Congruent (*M*=77%, *SE*=1.97), relative to Incongruent (*M*=70%, *SE*=1.79), responses. Thirdly, there was a large main effect of Premise type, *F*(1,33)=29.36, *p*<.001, showing that accuracy levels for End premises (*M*=83%, *SE*=1.99) were higher relative to Central premises, (*M*=64%, *SE*=2.36) highlighting a pattern of performance strongly indicative of a serial position effect. Indeed, in the incongruent condition central premise accuracy levels were not statistically different from chance levels (50%) after a Delay, as indicated by a one-sample t-test, *t*<1.

However, these main effects were not qualified by interactions between Delay type and Premise type, *F*<1, Delay type and Congruency, *F*(1,33)=1.84, *p*=.185, Congruency and Premise type, *F*(1,33)=3.07, *p*=.089, or three-way interaction between these factors, *F*<1. Consequently, although the serial position effect appears to be present in this analysis, it was not differentially expressed across the 24 hour delay.

Table 2. Experiment 1: Accuracy levels for central and end premises.

	Immediate condition		Delayed condition	
Congruency	End Premises (%)	Central Premises (%)	End Premises (%)	Central Premises (%)
Congruent	87 (4.6)	73 (4.2)	78 (5)	69 (5)
Incongruent	91 (4.7)	62 (3.9)	74 (5.1)	53 (4.2)

Note: End premise accuracy refers to A-B and E-F premise pairs. Central premise accuracy refers to C>D, D>E and E>F premise pairs. Values in parentheses denote the standard error.

RT performance and the serial position effect

Analysis of RT performance, which can be found in Table 3, did not reveal a main effect of Delay type, *F*(1,33)=1.78, *p*=.191, nor was there a main effect of Congruency, *F*(1, 33)=1.26, *p*=.296, despite the large impact of this factor on the previous analysis of accuracy performance. However, a strong main effect of Premise type did emerge, *F*(1,33)=29.7, *p*<.001. As with the above analysis of accuracy levels, this finding was in a direction indicative of a serial position effect, with faster RTs for End (*M*=1369, *SE*=72.7) relative to Central premises (*M*=1585, *SE*=79). Importantly, this main effect was qualified with a Delay type x List interaction, *F*(1,33)=5.9, *p*=.021, η_p^2=.152. This interaction, which can be seen in Figure 14, was driven by faster RTs for central, relative to end, premises after a 24 hour delay, a difference that was not expressed in the Immediate condition. No additional interactions were found between Delay type x Congruency, *F*<1, Congruency x Premise type, *F*(1,33)=2.49, *p*=.124, or three-way interaction between these factors, *F*<1.

Table 3. Experiment 1: RT performance for central and end premises.

	Immediate condition		Delayed condition	
Congruency	End Premises (msec)	Central Premises (msec)	End Premises (msec)	Central Premises (msec)
Congruent	1303 (94)	1763 (121)	1269 (103)	1448 (102)
Incongruent	1534 (140)	1699 (111)	1370 (152)	1431 (121)

Note: End premises refers to RTs for A-B and E-F premise pairs. Central premises refers to RTs for C>D, D>E and E>F premise pairs. Values in parentheses denote the standard error.

Taken together analysis of the central and end premises indicate the predicted 'flattening' of the serial position in the delayed condition. However this effect was only present for RT measures, as evidenced by a ~300 msec acceleration of RTs for central premises relative to end items after 24 hours. Interestingly, and counter to predictions, this effect does not appear to be differentiated by congruency, as would have otherwise been registered by a three-way interaction between the above factors.

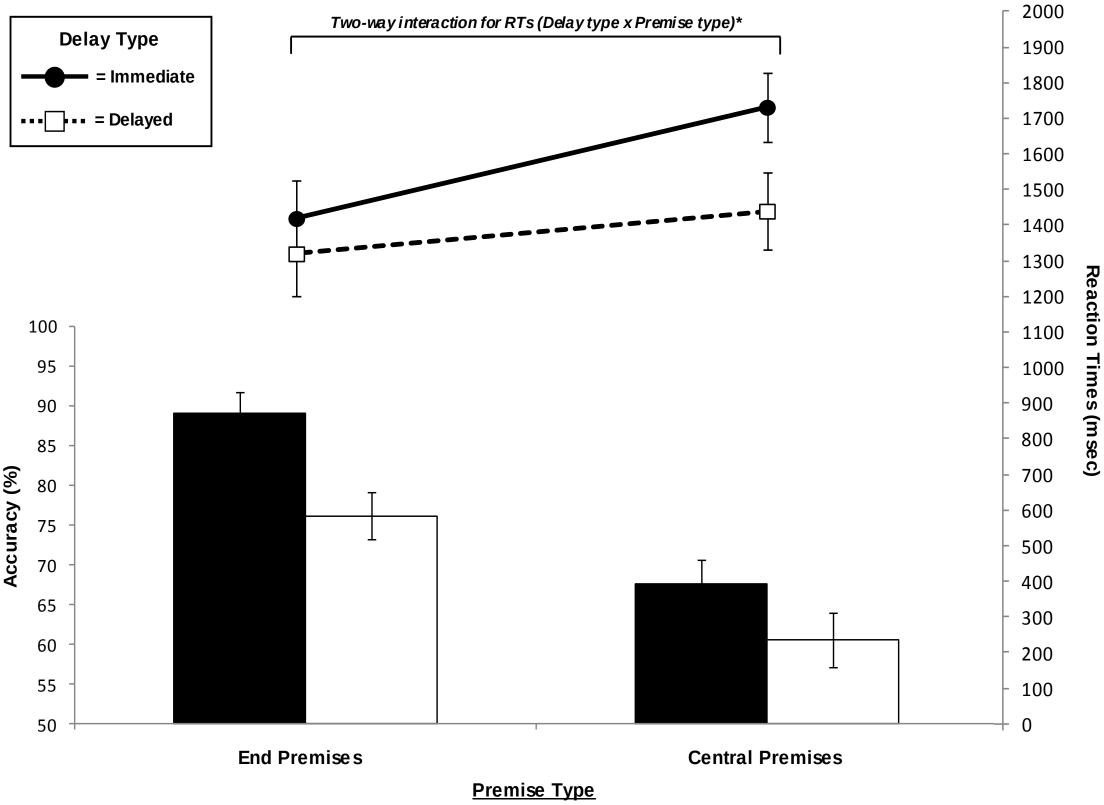


Figure 14. Experiment 1: Accuracy and RT performance between Delay type and Premise type. Significance value (*) indicates *p*<.05. Error bars represent standard error.

2.3.3 Transitive inferences and memory abstraction

The second integratory measure featured in this experiment relates to TI performance. Four novel comparisons presented at test were of interest to the present analyses: the strongly reinforced non-inference end item comparisons (A>F), novel word one unit transitive inferences (B>D), real word one unit transitive inferences (C>E) and two unit transitive inferences (B>E). These comparisons collectively formed the factor of Inference type which was entered into a 2x2x4 mixed-effects ANOVA featuring the between-participant variable of Delay type (Immediate vs. Delayed) and the within-participant variables of congruency (Congruent vs. Incongruent) and Inference Type (A>F, B>D, C>E and B>E).

The issue of training conformity

Before discussing the present analyses a brief comment should be made regarding the accuracy levels in the incongruent condition which, as can be seen in Table 4, were significantly below chance (as indicated by the symbol '•') excluding the non-inference (A>F) control comparisons. It is important to note that these lower accuracy levels do not necessarily represent 'incorrect' responses, particularly given the conflicting nature of the Incongruent premises underlying these inferences. Instead this observation suggests that when participants are dealing with Incongruent inferences they have shifted to a response pattern that consistently contradicts the associative order of trained information. As such, although congruent and incongruent responses will be compared in the present analyses, the dependent variable of 'accuracy (%)' will be relabelled as 'training conformity (%)' to highlight this differential pattern of performance. Moreover, due to the below chance levels of training conformity, RTs (as opposed to accurate RTs) will be used as a second dependent variable in these analyses.

Table 4. Experiment 1: Training conformity levels (%) for transitive inferences.

		Inference type			
Delay type	Congruency	Non-inference (A>F)	Novel word TI (B>D)	Real word TI (C>E)	Two unit TI (B>E)
Immediate	Congruent	100% (1.3)	75% (8.5)	91% (4)	89% (7.4)
	Incongruent	70% (8.7)	26% • (7.9)	5% • (4.1)	7% • (4.1)
Delayed	Congruent	97% (1.4)	74% (9.2)	92% (4.3)	73% (8.1)
	Incongruent	69% (9.4)	27% • (8.6)	9% • (4.5)	14% • (5.1)

Note: Values in parentheses indicate the standard error. The symbol ' • ' denotes levels of training conformity (%) that are significantly below chance levels (<50%) (see text).

Transitive inferences and training conformity

Mauchly's test indicated that the assumptions of sphericity were violated across Inference types, $\chi_2(5) = 21.05$ $p = .001$, $\varepsilon = .754$, and are therefore reported using a Greenhouse Giesser correction where appropriate. The results of training conformity analysis, which can be found in Table 4, indicated no main effect of Delay type, $F<1$, thus revealing little to distinguish the levels of training conformity over time. However, a large main effect of Congruency was evident, $F(1,33)=125.39$, $p<.001$, highlighting a profound Impact of the Congruency manipulation on performance, with higher levels of training conformity for Congruent (M=86%, SE=3.1) as opposed to Incongruent (M=28%, SE=2.88) responses.

Closer inspection of these data indicate that while all congruent responses were above chance (i.e. >50%) incongruent responses for both Delay types were below chance (<50%), while the non-inference (A>F) control comparisons remained relatively preserved at above chance levels. This was confirmed using a series of one sample t-tests revealing below chance levels of training conformity in the Immediate condition for B>D, $t(18)=-3.07$, $p=.007$, C>E, $t(18)=-14.57$, $p<.001$, and B>E, $t(18)=-16.73$, $p<.001$, inferences. Similarly, responses were below chance in the Delayed condition for the B>D, $t(15)=-2.7$, $p=.016$, C>E, $t(15)=-7.34$, $p<.001$, and B>E, $t(15)=-5.28$, $p<.001$, inferences. However, the non-inference (A>F) control comparisons remained significantly above chance in the Immediate condition, $t(18)=2.3$, $p=.034$, and marginally above chance in the Delayed condition, $t(15)=2.3$, $p=.061$, driving an overall main effect of Inference type, $F(2.26, 74.68)=36.76$, $p<.001$.

In terms of the critical interactions of interest to the present analysis, Delay type did not influence Inference type, $F<1$, nor was there an interaction between Delay type and Congruency, $F<1$, indicating that a 24 hour delay did not differentially impact on the levels of training conformity (%) for either of these factors. Moreover, no three-way interaction was evident between Delay type, Inference type or Congruency, $F<1$. Collectively, these results indicate that despite the notable influence of the congruency manipulation there was little to distinguish between the overall levels of training conformity across the 24 hour time delay.

Transitive inferences and RT performance

RTs were entered into the same 2x2x4 mixed-effects ANOVA as above. No main effect of Delay type was evident, $F(1,33)=2.72$, $p=.141$, indicating that RTs remained equivalent across time. Despite the large main effect of Congruency observed for the above training conformity measures, this main effect was not expressed in the RT measures, $F<1$. Importantly, a main effect of Inference type was present, $F(3,99)=15.33$, $p<.001$, which was further qualified by a highly significant interaction between Delay type and Inference type, $F(3,99)=8.3$, $p<.001$, $\eta_p^2=.362$ (discussed below). It should also be noted that an interaction between Congruency and Inference type was also present, $F(3,99)=10.7$, $p<.001$, however as this interaction is not of interest to the present analysis it will not be discussed further.

Table 5. Experiment 1: RT performance (msec) for transitive inferences.

		Inference type			
Delay type	Congruency	Non-inference (A>F)	Novel word TI (B>D)	Real word TI (C>E)	Two unit TI (B>E)
Immediate	Congruent	1123 (84)	2052 (127)	1398 (73)	1671 (109)
	Incongruent	1553 (146)	1849 (129)	1392 (94)	1394 (114)
Delayed	Congruent	1216 (91)	1492 (94)	1327 (80)	1394 (114)
	Incongruent	1553 (159)	1406 (140)	1312 (103)	1283 (125)

Note: All RT values are reported in milliseconds (msec). Values in parentheses indicate the standard error.

The critical interaction between Congruency and Inference type, which is illustrated in Figure 15, was driven by visibly faster performance for novel item TIs after a 24 hour delay. This was confirmed through a series of t-tests conducted between the two delay types. These analyses, which did not account for Congruency, indicated that RT performance across the time delay was not significantly different between the A>F end comparisons, *t*<1, the C>E real word TIs, *t*<1 , or the B>E two unit TIs, *t*(33)=1.6, *p*=.120. Importantly, however, the critical novel word inferences were found to be significantly faster after a 24 hour delay, *t*(33)=2.92, *p*=.006 (*mean difference* = 501 msec, *SE*= 171). Although this finding is in line with the predictions made for this measure, no three-way interaction was evident between Delay type, Congruency and Premise type, *F*<1. Consequently, while these measures indicate a clear acceleration of novel word TIs over time, this effect does not appear to be influenced by Congruency, despite the large impact of this manipulation on the underlying levels of training conformity.

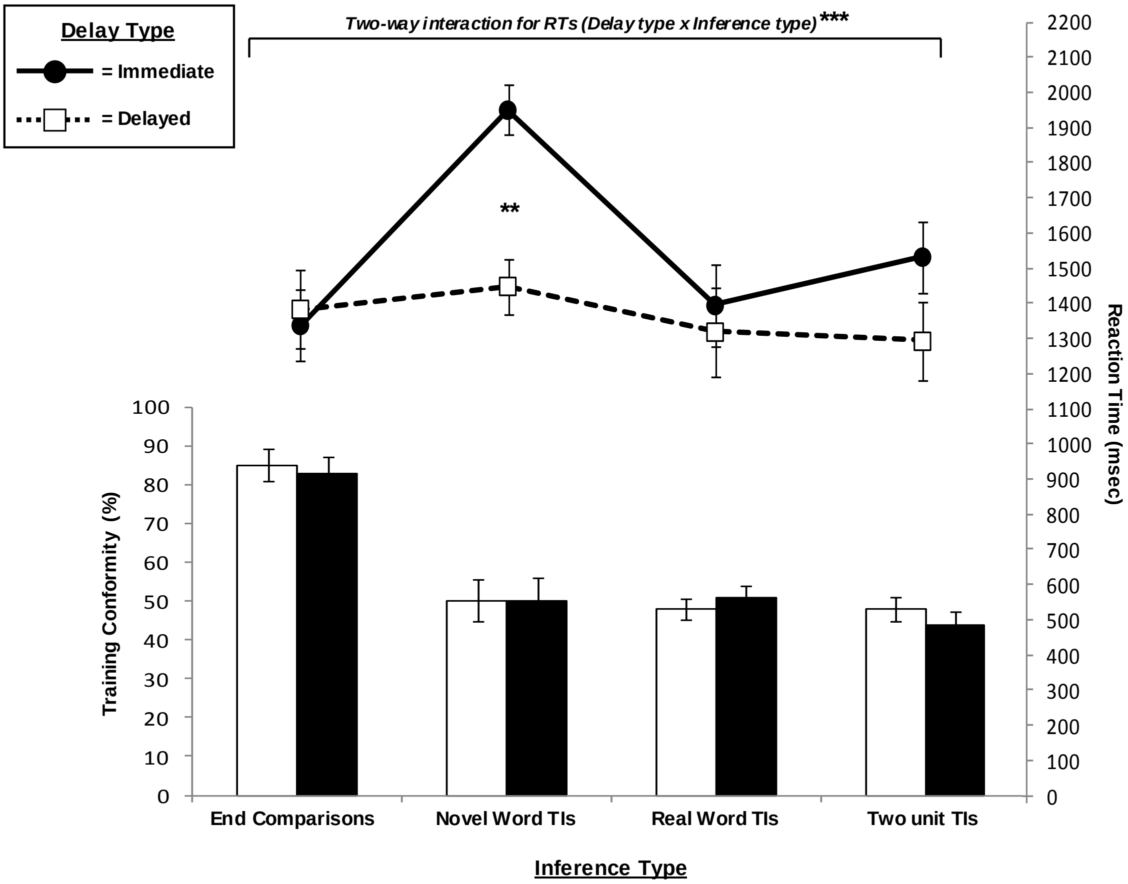


Figure 15. Experiment 1: Delay type and Inference type RT (msec) and Training conformity (%) performance collapsed across Congruency. Significance value (***) indicates *p*<.001, while (**) indicates *p*<.01. Error bars represent the standard error. Note that the training conformity values in this figure are collapsed across Congruency for both Delay types.

Given the impact of Congruency on training conformity RT analysis was repeated independently for Congruent and Incongruent RTs using the same ANOVA as above. For

Congruent RTs there was a significant Inference type X Delay type interaction, *F*(3,99)=7.86, *p*=<.001, driven by RTs in the Delayed condition for B>D responses that were 560 ms (*SED* =187) faster than those found in the Immediate condition. In the Incongruent condition the Inference type x Delay type interaction was marginally significant, *F*(3,99)= 2.57, *p*=.059, with RTs for the B>D inferences faster by 443 ms (*SED* = 190). As such, the speeding of B>D RTs over 24 hours appears to be reflected in both conditions, albeit more strongly for exclusively Congruent responses.

2.3.3 Size generalisations and the semantic distance effect

The third and final measure of integration in this experiment relates to the semantic distance effect for size judgements between novel words and new, but familiar, real word items. As established in section 2.1.2, the expression of the semantic distance effect on this measure should be reflected in faster responses for comparisons featuring larger size distance. To explore this possibility a factor of Proximity was created by allocating size comparisons within the closest three size bands (0-3) of a given novel word as Smaller, and comparisons involving the three most distant size bands (3-5) as Larger, respectively. This factor was entered into 2x2x2 mixed-effects ANOVA featuring the between-participant factor of Delay type (Immediate vs. Delayed) and the within-participant factors of Congruency (Congruent vs. Incongruent) and Proximity (Smaller vs. Larger) using the dependent variable of RT performance.

The products of this analysis, which can be found in Table 6, did not reveal a main effect of Proximity, *F*(1,33)=1.33, *p*=.256, or Congruency, *F*<1, however a marginal main effect of Delay type was present, *F*(1, 33)=2.9, *p*=.098, with participants in the Delayed condition (*M*=1477, *SE*=89) showing a tendency for overall faster responses on this measure than individuals in the Immediate condition (*M*=1684, *SE*=82). However, this finding was not qualified with an interaction between Delay type and Proximity, *F*<1, or a three-way interaction between Delay type, Congruency and Proximity, *F*<1. Therefore, aside from the tendency for decreased RT levels across 24 hours, there was little evidence for the predicted influence of Delay type on the expression semantic distance effect.

Table 6. Experiment 1: RTs and the size proximity of generalisations for congruent and incongruent novel words.

	Immediate condition		Delayed condition	
Congruency	Smaller proximity (msec)	Larger proximity (msec)	Smaller proximity (msec)	Larger proximity (msec)
Congruent	1718 (94)	1614 (98)	1584 (102)	1423 (107)
Incongruent	1696 (88)	1709 (106)	1414 (96)	1489 (115)

Note: Proximity refers to smaller (0-2) or larger (3-6) size band distances between a given novel word and new, but familiar real word items. Values in parentheses denote the standard error.

Despite the lack of differences relating to the time delay manipulation a Congruency x Proximity interaction was present, *F*(1, 33)=4.24, *p*=.048, which is illustrated in Figure 16. This interaction was explored further using paired sample t-tests between Congruency and within the size Proximity factors. In line with a potential semantic distance effect, it was found that size judgements in the Congruent condition were faster for Smaller (*M*=1651, *SE*=69.37) relative to Larger distances (*M*=1519, *SE*=72.28), *t*(34)=2.13, *p*=.04, while no such difference was observed in the Incongruent condition, *t*<1. As such, while the size proximity effect was not differentiated by Delay type, there is evidence that the semantic distance effect was present for Congruent, but not Incongruent, novel words size generalisations.

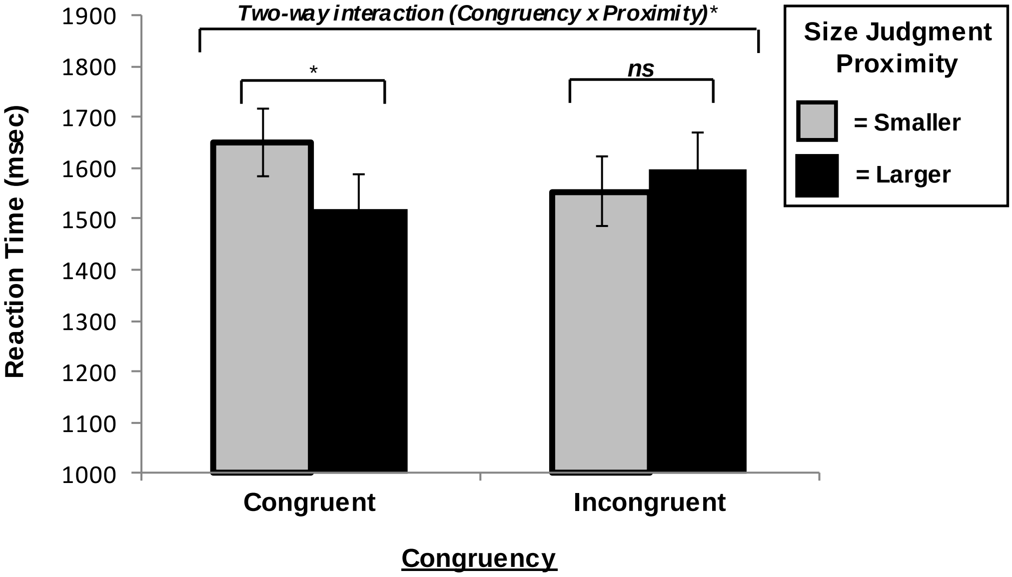


Figure 16. Experiment 1: Congruent and Incongruent size judgment distances aggregated across Delay type. Significance level (*) indicates *p*<.05, ns = not significant. Error bars represent standard error.

2.4 Summary and Discussion

Previous studies implementing the TI paradigm have revealed consolidation-associated enhancements to memory (Ellenbogen et al., 2007; Werchan & Gómez, 2013). However, the underlying premises featured in these investigations shared little or no relationship with prior knowledge, a factor otherwise considered important for mnemonic integration in current theoretical (McClelland et al., 1995; Stickgold & Walker, 2013) and experimental accounts of consolidation (Tse et al., 2007; van Kesteren, Fernández, et al., 2010; van Kesteren, Rijpkema, Ruiter, & Fernández, 2010). The present experiment was designed to build on this literature by exploring how the semantic congruency of newly learned information might influence its subsequent integration 20 minutes, or 24 hours, after training. In line with previous research it was predicted that congruent information would display more signs of enhanced integration on three behavioural measures after an extended delay.

Several key findings emerged from this investigation. Firstly, a consolidation associated suppression of the serial position effect was revealed, as evidenced in the speeding of RTs for central premises over 24 hours. Secondly, responses for novel word TIs were also faster after an extended delay, while no such difference emerged for the non-inference or semantic inference control comparisons. These differences in performance are not accountable through other potential factors such as baseline differences in sleep quality, training performance or levels of alertness. When considered together these findings align with the overarching prediction that consolidation will facilitate the integration of new information over time. Importantly, however, these effects were not differentiated by congruency as originally hypothesised. This observation is particularly interesting given the substantial impact of the incongruent condition on the levels of accuracy underlying these measures. Despite these seemingly contradictory aspects of the present dataset, these findings can be reconciled with the notion that relevant prior knowledge can facilitate the ongoing consolidation and integration of declarative memory.

2.4.1 The issue of congruency and consolidation

Before discussing the specific findings of the present experiment the overall impact of congruency needs to be addressed. Specifically, the overarching hypothesis framing this investigation predicted a differential impact of congruency on integration. As noted above, although this manipulation did strongly influence the levels of accuracy and training conformity, congruency itself did not significantly impact the principle markers of integration in terms of RT performance. In resolving this apparent dichotomy two factors need to be considered: (i) the impact of prior knowledge and expectations of participants when interpreting incongruent information, and (ii) crossover effects regarding the congruency manipulation itself.

The simplest explanation for the present findings is that participants were using their pre-existing knowledge to re-interpret incongruous information as congruent prior to a consolidation opportunity. Here participants may have ‘ignored’ the incongruent component of training in favour of their own congruous ordering of stimuli relations. This is a distinct possibility given that incongruent premises were introduced covertly during the training phase, with no instructions provided on to how to deal with this conflicting information when it arose. This appears to have occurred soon after training (within 20 minutes), and was most prominently expressed in the below levels of performance on the TI measures, suggesting that participants may have applied their own semantically consistent dimension of size to otherwise nonsensical information.

A second possibility is that congruent training in the present experiment may have provided participants with cues on how to interpret incongruent information, particularly given that these conditions were trained in close proximity. Previous experiments relating to consolidation and prior-knowledge contain protocols that largely preclude an influence of one congruency condition on the other. For example, in the Tse et al. (2007) investigation, rats were trained on congruent or incongruent food-reward locations in separate and highly distinct arenas, thus negating any practical application of congruent learning beyond its original trained context. Moreover, the scrambled (incongruent) or unscrambled (congruent) film clips featured in the van Kesteren and Fernández, et al. (2010) investigation were part of a between-participant design, similarly removing the opportunity for one condition to impact performance on the other.

However, in the present experiment exposure to the logically consistent associative and semantic ordering of congruent information may have influenced the way participants interpreted, and subsequently consolidated, incongruent information. This factor may have contributed to the progressive re-interpretation of incongruent information based on experience. Indeed, relevant to this notion are previous studies showing that consolidation is particularly beneficial in the extrapolation and application of learned rules to new contexts and situations. This has been demonstrated on a variety of measures, including the ability to apply novel grammatical rules to new combinations of auditory information (Gómez et al., 2006) the extrapolation of statistical frequencies in a tone learning task (Durrant et al., 2011) and explicit insight into hidden rules underlying a number reduction task (Wagner et al., 2004).

Importantly, the above accounts are not mutually exclusive, and can simultaneously accommodate the large impact of congruency and concurrent patterns of integration regarding RT performance in the present experiment. If incongruent information is encoded and consolidated in a ‘congruent-like’ form, then diminished accuracy levels will inevitably arise, as incongruent responses increasingly diverge from the trained associative order of incongruent information. Integration, as indicated by the RT measures, can still occur as low accuracy levels do not necessarily reflect inaccurate responding *per se*, but rather a diverging set of response choices that reflect a more coherent reinterpretation of the incongruent information.

2.4.2 Consolidation facilitates the suppression of the serial position effect

The first of the three integrative measures in this experiment explored the influence of consolidation on the serial position effect, an often-expressed behavioural phenomenon derived from the stronger reinforcement history of end premises relative to central premises (Acuna et al., 2002; Greene et al., 2001; Moses et al., 2006; Ryan et al., 2009). The prediction that consolidation would enhance performance in a direction counter to this effect was validated by the ~300 msec speed increase for premises central to the schematic hierarchy after a 24 hour delay. An equivalent speeding of RTs was not observed for the adjacent end premises, measures which provide a useful gauge of baseline performance in relation to the strong reinforcement history of the end premises (Frank et al., 2005; von Fersen et al., 1991). While these findings are in a direction indicative of consolidation-associated integration, the manipulation of congruency itself did not interact with the above noted effect, despite the large drop in accuracy for incongruent information, both immediately, and after an extended delay.

It is important to note that the present findings are not easily accountable in terms of a speed/accuracy trade-off, a factor that needs to be considered given the low levels of incongruent accuracy observed after a 24 hour delay. Firstly, counter to this possibility, the accuracy levels associated with the central premises remained equivalent between the two time delays, and it was on these particular premises which an acceleration of RTs was observed. Secondly, if a speed/accuracy trade-off was indeed present we might expect a global speeding of RTs in relation to the lower accuracy levels seen in the incongruent condition, however there was no main effect of time delay on these measures. Thirdly, unreported analysis restricted to the congruent condition showed an almost identical pattern of results, with no differences between the Immediate and Delayed conditions in terms of accuracy (~80%) but faster accurate RTs (by approximately ~370 msec) for central premises. As such, in terms of RT performance, the present findings strongly suggest a legitimate influence of time, and potentially consolidation, with regards to the suppression of the serial position effect.

The question remains as to the nature of the integration underlying this finding. In the predictions made for this measure two consolidation-associated factors were hypothesised to counteract the serial position effect. Firstly, it was proposed the assimilation of premises into pre-existing knowledge might ‘extend’ the otherwise finite size limits framing the schematic dimension, effectively reducing the impact of the strongly reinforced premises marking the end of this hierarchy. A second proposition was that the shared reactivation of overlapping associative and semantic features during sleep may have increasingly bound together premise pair information, resulting in a more unified representation of these items in memory (Lewis & Durrant, 2011; Walker & Stickgold, 2010). While these accounts are not mutually exclusive, the present results provide more support for the latter ‘unitisation’ explanation. Specifically, there was no evidence for a slowing of end premise RTs in terms of the serial position effect, which might be predicted if memory assimilation was weakening the influence of these strongly reinforced items.

Instead, RT acceleration was restricted to central premises, more in line with the unitisation account, particularly given that RT levels across the entire schematic dimension were rendered increasingly equivalent after 24 hours (Kuriyama et al., 2004).

An additional possibility is that consolidation may be facilitating the integration and familiarisation of novel word forms over time, an account that can be considered distinct from the semantic integration proposed here (Leach & Samuel, 2007). For example, Tamminen, et al. (2010) found that responses in the catagorisation of novel words in an old/new recognition task were accelerated by ~200 msec after a consolidation opportunity. However, the large effect sizes seen in the present experiment (>300 msec), particularly in relation to the measures of TI discussed in the following section (~500 msec), suggest that the consolidation-associated enhancement of novel word forms alone is unlikely to be the only contributing factor to the present findings.

2.4.3 Consolidation facilitates the development of relational memory

The second of the three integrative measures featured in this experiment related to the predicted enhancement of TI performance over 24 hours, particularly in relation to inferences formed exclusively between two pieces of novel information. This hypothesis was confirmed through a striking ~500 millisecond speed increase for novel word TIs (B>D) after 24 hours. Importantly, RT levels did not differ across the time delay with regards to the strongly reinforced non-inference comparisons (A>F), or inferences formed between already known word items (C>D). This finding suggests that the presence of relevant semantic information during training can help imbue completely novel information with enhanced relational properties over time.

While congruent inferences are neatly accommodated by the current experimental hypothesis the observations regarding incongruent inferences are not. Here, a pattern of integration similar to the congruent condition arose, as manifested through an equivalent speed enhancement for novel inference items. Furthermore, training conformity levels, excluding the non-inference comparisons, were all far below chance for incongruent inferences, indicating successful TIs that contradicted the original training. This may initially appear somewhat counterintuitive, particularly given that classic accounts of TI performance propose that inferences are directly deduced from the memories of the underlying premises (Bryant & Trabasso, 1971; Piaget, 1928). Here, despite the successful capacity to form incongruent TIs at both time delays, accuracy levels for incongruent premises were low 20 minutes after training, and were at chance levels after 24 hours.

As with performance on the underlying premises, incongruent TIs appear to have been strongly influenced by the way participants have applied their pre-existing knowledge shortly after training. This is most apparent on the TIs that involve size judgements between two already known items. For example, the below chance levels of training conformity on these incongruent inferences demonstrate that a logical semantic relationship (e.g. 'giraffe>lizard') was favoured over the counterintuitive responses required had participants followed the associative order of incongruent training (e.g. 'lizard>giraffe'). While this observation is perhaps unsurprising, what is interesting is that all of the incongruent TIs follow suit, similarly displaying below chance levels of training conformity after 24 hours. As such, participants appear to be judging all incongruent TIs in a 'congruent' manner, even when these comparisons featured entirely novel information. As this effect was present at both time delays, there is good evidence that incongruent information was 'reinterpreted' in a more congruent form prior to the 24 hour delay, and consolidated as such over the subsequent 24 hours.

Moreover, the present observations suggest that inferences were being derived from an overarching size dimension rather than deduced directly from the memory of the trained premises, as classic models of TI performance have previously proposed (Bryant & Trabasso, 1971; Piaget, 1928). Relevant to this observation is the 'gist' account of TI performance, where it is argued that participants will extrapolate features across premises to form inferences (i.e. gist), rather than establishing these inferences directly from premises (i.e. veridical memory) (Brainerd & Reyna, 2002; Healy, Kosslyn, & Shiffrin, 1992)(Reyna & Brainerd, 1990). Here, gist memory is thought to help maximise our predictive behaviour in relation to experience while simultaneously reducing the need to retain the episodic details underling these experiences. In terms of consolidation, the extrapolation of gist is argued to account for findings in the false memory literature, namely the observation that when participants are trained on themed word lists (e.g. 'bed', 'snooze', 'pillow') they will selectively retain false memories of word items maximally associated with this semantic dimension (e.g. 'sleep') while memory of the original word items simultaneously diminishes after a period of consolidation (Diekelmann et al., 2010; Payne et al., 2009).

A gist account accommodates several features of the present dataset. Firstly, as above, it explains why participants are capable of performing incongruent TIs despite the low levels of accuracy underlying these premises. Secondly, the extrapolation of gist is thought to be particularly useful when inferring relationships across incomplete or inconsistent information (Reyna & Brainerd, 1990), perhaps explaining the parallel integration observed across the congruency manipulation. Thirdly, a 'congruent' gist-like rule may have also influenced participant's memory of the underlying premises. The rapid drop in accuracy for incongruent premises 20 minutes after training, which became more pronounced after a consolidation opportunity, may reflect an increasing tendency for incongruent responses to become more congruent (and thus less 'correct') on this measure over 24 hours.

2.4.4 Diverging routes of consolidation

The present findings contribute to the small but growing body of literature indicating that consolidation can promote relational memory (Coutanche, Gianessi, Chanales, Willison, & Thompson-Schill, 2013; Lau et al., 2011, 2010). However, counter to previous investigations using the TI paradigm (Ellenbogen et al., 2007; Werchan & Gómez, 2013), no consolidation-associated differences were found for the more distant two unit TIs (i.e. B>E inferences). Comparisons between the present experiment and these previous investigations must be made tentatively as the current findings relate predominantly to changes in RT performance which have not been reported previously (Ellenbogen et al., 2007; Werchan & Gómez, 2013). Nevertheless, the lack of notable enhancement for two unit inferences in the present investigation, both in terms of training conformity (%) and RTs, remains intriguing.

The mechanisms underlying generalisation in previous TI experiments, as established by Werchan and Gómez (2013), appear to derive from a form of procedural consolidation. Inferences of this type are thought to be attributable, at least in part, to a lack of explicit awareness of the overarching stimuli relations (Greene et al., 2001; Greene, 2007), with associative reinforcement generalising responses as a 'fallback' mechanism in the absence of explicit declarative knowledge (Frank et al., 2006, 2005). In the present experiment, new information was interleaved with relevant semantic information, potentially increasing the applicability of this information to future behaviour (McClelland & Goddard, 1996; McClelland & Rogers, 2003; O'Reilly & Norman, 2002). However, in the absence of such information, procedural consolidation may still permit an adaptive degree of generalisation (Frank et al., 2006, 2005). In accordance with this position, the consolidation-associated effects noted by Ellenbogen et al. (2007) and Werchan and Gómez (2013) did not necessarily require any superordinate knowledge of relationships across the schematic hierarchy, whereas the consolidation-associated effects seen in the present investigation did.

Therefore, one potential explanation for the divergent findings across these investigations is that separate routes of consolidation are in operation, each geared towards maximising the future utility of memory (Stickgold & Walker, 2013). Relevant to this notion is an experiment conducted by Wagner, Gais, Haider, Verleger, and Born (2004) where participants were trained on a number reduction task that could be rapidly solved once insight into a hidden rule was discovered. Of the participants that slept, a majority (59%) developed insight relative to wakefulness. Interestingly, however, the remaining participants who failed to develop insight instead produced faster RTs across the trained sequences relative to individuals who had obtained insight or remained awake. In this case consolidation appears to have diverged down two potential routes, each conferring unique benefits dependent on the relative access to declarative insight (Stickgold & Walker, 2013). Similarly, the differential findings observed across the previous and present consolidation-associated TI experiments suggests two adaptive routes of consolidation which may derive, at least in part, from the presence or absence of semantic information.

2.4.5 Novel word generalisations

The third and final measure of integration in this experiment assessed the generalisation of newly learned novel words to untrained, but familiar, word items. Generalisation was assessed in relation to the semantic distance effect which, if present, would be registered through slower responses between size judgements featuring items closer together in size (Pavio, 1975; Rubinsten, Henik, Berger, & Shahar-Shalev, 2002). This effect is thought to be contingent on the saliency of mental comparisons along an internal dimension of size. Accordingly, it was predicted that that the assimilation of novel words into semantic memory over 24 hours might heighten the relative imagability of these items, thus inducing a stronger semantic distance effect.

Although this measure did reveal a semantic distance effect for congruent size comparisons, as expressed in the aforementioned slowing of size judgements closer together in size, the strength of this effect did not differ across time. Although this finding reflects a null result it should also be noted that there was a marginally significant trend for faster responses on this measure after 24 hours. It may be that the higher order assimilation of novel words requires an extended period of consolidation, potentially involving multiple nights of sleep, as noted in other investigations exploring the time course of memory consolidation (Meeter & Murre, 2004; Tamminen et al., 2010; Walker, Brakefield, Seidman, et al., 2003).

That said several drawbacks inherent in this measure limit the extent that firm conclusions can be made. Firstly, there was no baseline comparison measure (e.g. real words vs. real words) making it hard to distinguish the extent of integration in relation to an already established semantic distance effect. Moreover, the present results do not preclude some form of generalisation in the incongruent condition. As shown, the evidence suggests participants were forming incongruent inferences on the basis of their own internally consistent re-interpretation of incongruent information. It is not unreasonable to assume that this factor may have also influenced size generalisations beyond the trained context. This possibility, unfortunately, would not be registered by this measure in its current form.

Experiment 2, presented in the following chapter, addresses some the noted limitations of this measure.

2.4.6 Conclusions

Using a modified TI paradigm the present findings show that the presence of semantic information during schematic training is sufficient to induce rapid mnemonic integration, as evidenced in the speeding of central premise responses and novel word TIs over 24 hours. These changes to behaviour are not accountable through differences in alertness, sleep patterns or baseline levels of performance. Moreover, the integration seen in this experiment differs from that seen previously using the TI paradigm. Here, the influences of consolidation no longer appear to be restricted to the reinforcement history underlying premises and are instead reflected behavioural changes across the entire schematic dimension. As these integrative effects were not dissociated by semantic congruency these findings implicate two important roles for prior knowledge: firstly in facilitating the extrapolation of relevant information prior to a consolidation-opportunity, and secondly, in structuring the ancillary integration of this information over 24 hours.

Chapter 3: Consolidation, sleep and memory integration

3.1 Introduction

Experiment 1 demonstrated that when new information relates to semantic memory it can show signs of rapid integration over 24 hours. Following on from this investigation, Experiment 2 aims to explore how sleep and its associated physiology may contribute to these behavioural outcomes. As reviewed in Chapter 1, there is now considerable evidence that sleep-associated memory consolidation can help optimize declarative memory for future use (Conte & Ficca, 2012; Lewis & Durrant, 2011; Walker & Stickgold, 2010) including the assimilation and abstraction of new information into predictive semantic networks (McClelland & Rogers, 2003; Rogers & McClelland, 2004). The integration of new memories into such a richly interconnected system may imbue them with enhanced relational properties, as suggested in both the existing literature (Ellenbogen et al., 2007; Lau et al., 2010; Payne et al., 2009; Tse et al., 2007) and the investigation conducted in the previous chapter.

On this basis the present experiment aims to explore the potential mechanistic contributions of sleep physiology on memory integration using the modified TI paradigm featured in Experiment 1. Current models of sleep-associated memory consolidation relate to various aspects of the sleep state; both at the level of sleep stage activity and the numerous field potentials underlying these sleep stages. The rest of this introduction will therefore provide a short overview of these accounts. Although these models were previously introduced in Chapter 1, here they are discussed in relation to the potential outcomes of this investigation.

3.1.1 Memory integration and sleep stages

In Chapter 1 it was shown that the sleep state consists of a number of physiologically and neurologically distinct sub-stages including rapid eye movement (REM) sleep, and sleep stages 1 through 4, which correspond to the increasing ‘depth’ of sleep activity, collectively referred to as non-rapid eye movement (NREM) sleep (Rechtschaffen & Kales, 1968). Moreover, the deepest stages of NREM sleep (stages 3 and 4) are together known as slow wave sleep (SWS). All of these sleep stages, with the exception of stage 1 (which typically marks a short transitionary period from wakefulness to sleep), have been affiliated in some way with declarative memory consolidation, and thus may contribute to the mnemonic integration observed in Experiment 1 (Stickgold, 2009).

REM sleep and integration

While REM sleep has traditionally been associated with the consolidation of procedural memory, as typically measured using serial reaction time and visual motor tasks

(Karni et al., 1994; Mandai, Guerrien, Sockeel, Dujardin, & Leconte, 1989; Plihal & Born, 1997), there is now good evidence that this state may also play a role in the consolidation of declarative memories, particularly for those with emotional content (Fogel et al., 2007; Goerke et al., 2012; Wagner et al., 2001, 2006). More recently, REM sleep has been co-opted as an ideal state for integrative consolidation, permitting the unitization (binding), abstraction and assimilation of new information into existing memory (Walker & Stickgold, 2010)(see Chapter 1, Figure 6). Neurologically speaking, integration in this manner is proposed to benefit from the heightened cortico-cortico connectivity present during REM sleep relative to other sleep stages (Braun et al., 1997) which may support associative linking between information held in disparate cortical regions (Fogel et al., 2007; Jones & Wilson, 2005).

As each of the integrative processes outlined in this model (unitization, abstraction and assimilation) are conceptually related to the three measures of integration in the present experiment (premise pair performance, TIs and size generalisations) we may therefore expect to see a corresponding relationship with REM sleep. It should be noted, however, that while unitization, assimilation and abstraction have all been shown to benefit from sleep (Ellenbogen et al., 2007; Kuriyama et al., 2004; Tse et al., 2007), there is presently little direct evidence linking these shifts in memory expression to REM sleep itself (Diekelmann & Born, 2010b).

NREM sleep and integration

Elsewhere in the literature, particular attention has been drawn to slow wave oscillations (<1 Hz) that predominate SWS activity in two non-mutually exclusive accounts of consolidation. Firstly, the synaptic homeostasis hypothesis predicts that the oscillating up-states and down-states in SWS may help globally reduce the expression of superfluous synaptic connections, thus enhancing the signal to noise ratio of memory traces robust enough to survive this downscaling process (Tononi & Cirelli, 2003, 2006). Although this account has received some support (Huber et al., 2006, 2004; Mednick et al., 2002), in Chapter 1 it was noted that this model by itself cannot accommodate the repeated observation that sleep can enhance relatively impoverished declarative memories (Diekelmann et al., 2010; Drosopoulos, Schulze, et al., 2007; Schmidt et al., 2006). Consequently, should such homeostatic processes exist, they must be accompanied by more selective forms of consolidation that are not completely contingent on post-encoding memory strength (Stickgold & Walker, 2013).

A second mechanistic account of SWS activity, the active consolidation model, proposes that slow wave oscillations will facilitate the consecutive reactivation of recently acquired memories in the hippocampus with distributed elements of cortical knowledge, thus guiding the selective uptake and integration of new information (Buzsáki, 1996; Frankland & Bontempi, 2005; Gais et al., 2007; McClelland et al., 1995). More specifically, slow waves generated by the neocortex are thought to operate in synchrony with sharp wave ripples, high frequency (200 Hz) bursts of activity in the hippocampus, and sleep spindles, short bursts of 12-15 Hz activity in the thalamus. The coalescence of these three field potentials are thought to guide the bidirectional exchange of information between the hippocampus and neocortex, as slow oscillations trigger sharp wave ripples which in turn become nested in the single oscillatory troughs of spindles. These three factors are thought to drive mnemonic reactivation at a point when neuronal networks are optimally predisposed to persisting synaptic changes (Mölle et al., 2009, 2002; Mölle, Marshall, Gais, & Born, 2004).

In accordance with both the homeostatic and active consolidation accounts, the sleep stages most consistently associated with declarative memory consolidation relate to NREM sleep, and in particular SWS, as evidenced through various measures of word pair learning. In terms of quantitative shifts in memory performance NREM sleep has been found to correspond with the improved recall of unrelated word pairs (e.g. '*peach-chair*') both after a nap (Tucker & Fishbein, 2008) and a full night of sleep relative to wakefulness (Barrett & Ekstrand, 1972). Conversely the disruption of NREM sleep, but not REM sleep, can impaired the recall of unrelated word pairs (Ficca et al., 2000). Furthermore, SWS-specific activity has been associated with the consolidation of semantically related word pairs (e.g. '*river-boat*'), as shown through improved memory recall after periods of SWS rich nocturnal sleep (Plihal & Born, 1997, 1999), and after a brief nap daytime containing SWS (Tucker et al., 2006). Lastly, although the above studies are correlational in nature, causality for the active consolidation model is strongly supported through studies that cue of memory reactivation during NREM sleep. For example, when information is encoded in the presence of salient sounds or smells, covert re-exposure to these stimuli in NREM, but not REM, sleep has been shown to enhance memory performance on a range of declarative tasks (Bendor & Wilson, 2012; Rasch et al., 2007; Rudoy et al., 2009).

More recently, and of particular relevance to the present investigation, SWS has been associated with a number of more qualitative changes in the expression of declarative memory (Mazzoni et al., 1999; Smith, 2001). For example, SWS over a nap has been associated with improvements in the relational memory of arbitrarily linked elements of information (pictures of faces and objects), implicating this state in the flexible expression of declarative knowledge (Lau et al., 2010). Similarly, SWS has been associated with the ability to abstract predictive relationships in a probabilistic tone learning task (Durrant et al., 2011) as well as increased rates of false memory recall, which appear to be contingent on the abstraction of shared features across semantically themed word lists (Brainerd & Reyna, 2002; Payne et al., 2009). Given the clear association between SWS and the extrapolation of shared features across both related and unrelated declarative materials we may expect a similar relationship to emerge in the present experiment.

3.1.2 Memory integration and sleep spindles

Given that the active consolidation model relates to both sleep stage activity and the field potential oscillations underlying these stages, there is an increasing call for researchers to explore these latter aspects of sleep physiology (Conte & Ficca, 2012; Diekelmann & Born, 2010a). To this end there is evidence that sleep spindles during NREM sleep can provide a useful biomarker of mnemonic replay between the hippocampus and neocortex, and thus by extension potential memory integration (Buzsáki, 1996; Sirota & Buzsáki, 2005; Sirota et al., 2003).

Again, in terms of the more quantitative changes in the expression of declarative memory, sleep spindle activity has been associated with the enhancement and stabilization of recently acquired memories. For example, spindles have been associated with the retention and recall of unrelated word pairs that are both concrete (e.g. '*brick – tree*') (Clemens et al., 2005; Gais et al., 2002; Schabus et al., 2004) and abstract in nature (e.g. '*union – rate*') (Schmidt et al., 2006). Furthermore, while these studies are correlational in nature, a more causal relationship has recently been established through the administration of the drug zolpidem, which has been shown to greatly stimulate spindle activity, albeit while simultaneously truncating REM sleep. Here, it was found that increased spindle density in a zolpidem treatment group relative to a placebo was positively associated with the improved retention of semantically related word pairs (Mednick et al., 2013). However, in this latter case it should be noted that Zolpidem has substantial residual effects, and thus although it may have desirable effects on sleep, it is unlikely that cognitive performance measured the next day is entirely down to the influence of this drug alone (Groger 2013, personal communication).

In terms of memory integration, spindle activity during NREM sleep has been associated with the emergent lexical competition between novel word forms (e.g. '*cathedruke'*) and related pre-existing word forms (e.g. '*cathedral'*) (Tamminen et al., 2010). However, it is important to note that the Tamminen et al., (2010) investigation, unlike the present study, assessed the integration of word forms in the absence of any semantic features. Overall, it would appear that the use of sleep spindles as an index for consolidation is well justified, particularly if learning engages hippocampal structures, a region associated with relational memory performance in the TI paradigm (Heckers et al., 2004; Kumaran & McClelland, 2012; Van Elzakker, O'Reilly, & Rudy, 2003).

Given the prominent role of SWS in the standard model, we might expect spindles operating in these stages to have particular relevance to declarative memory consolidation. Indeed, recent evidence indicates that memory for contextual details are supported by SWS specific spindle activity (Cox et al., 2012) especially for information deemed to be of future relevance (Wilhelm et al., 2011). In the present experiment, training information as semantically congruent may similarly prime information to be of future utility, potentially triggering a 'fast-tracked' route of integration (Kroes & Fernández, 2012; van Kesteren et al., 2012). Given the previously noted evidence relating SWS to rule abstraction (Durrant et al., 2011) generalisation (Payne et al., 2009) and the development of relational memory (Lau et al., 2010) we should expect to see a prominent relationship between SWS spindle activity and memory integration in the present investigation.

3.1.3 Experiment aims and predictions

The present experiment implements the paradigm featured in Experiment 1 to focus on the relationship between sleep physiology and integrative consolidation, however the incongruent condition has been removed from this particular investigation for three reasons. Firstly, incongruency was minimised to obtain cleaner measures of potential integration to correlate with sleep physiology. There was evidence in Experiment 1 that the incongruent condition led to a degree of participant confusion regarding the paradoxical training, likely accounting for the considerable variability on some of the measures. Secondly, the nature of the incongruent condition made it difficult to interpret some of the integrative measures, specifically those relating to generalisation. Thirdly, the key findings of the previous investigation appeared to derive from the congruent interpretation of stimuli regardless of trained congruency. The present experiment therefore aims to focus on the nature of congruent integration driving this effect.

To this end, Experiment 2 explores two equivalent time delays involving either a period of sleep or wakefulness, as part of what is known as a '12:12' design. These conditions not only permit behavioural comparisons between a period of sleep relative to wakefulness, but also an exploration of the relationship between integration and the architecture of sleep itself. On the basis of the findings from Experiment 1, and the literature reviewed here, the sleep condition is predicted to manifest levels of integration not observed after a period of wakefulness. It should be noted that an 'immediate' 20 minute condition is also included in this experiment to mitigate the methodological shortfalls inherent in a pure 12:12 design (see Chapter 1). The inclusion of this condition will also permit a direct comparison with Experiment 1.

More specifically, the presence of congruent semantic information is predicted to facilitate performance on (i) central premises, as manifested through a suppression of the serial position effect, and; (ii) novel word TIs. In the former instance the suppression of the serial position effect is proposed to derive from the unitisation (binding) of the features shared across central premises as they become increasingly integrated into semantic memory (Kuriyama et al., 2004; Walker & Stickgold, 2010). In the latter instance RTs are predicted to decrease for novel word TIs due to their relative integration into pre-existing semantic networks, resulting in faster relational judgements between these elements of information (McClelland & Rogers, 2003; Reyna & Brainerd, 1990). Furthermore, integration on these measures is predicted to relate to aspects of sleep architecture, in particular SWS and its affiliated spindles.

The generalisation measures featured in Experiment 1 were also retained for this investigation, albeit with several modifications (outlined in the following section). While little was found to distinguish the two time delays on this measure in Experiment 1, there was nevertheless some evidence for a semantic distance effect featuring congruent novel vs. real word comparisons at both the immediate and delayed time points. Furthermore, RTs in the delayed condition were marginally (and globally) faster, potentially indicating the emergence of more efficient generalisations in this condition after 24 hours. Should sleep facilitate the relative assimilation of novel items into an overarching size dimension, as outlined in Chapter 2 (section 2.1.3), we may expect to see a relationship between sleep physiology and the positive expression of the semantic distance effect on this measure. Moreover, the influence of consolidation should be restricted to comparisons between novel and real word items, as opposed to a newly introduced baseline measure featuring comparisons between real words already existing in memory.

3.2 Methods

As the procedures used in the present experiment are almost identical to those featured in Experiment 1 only deviations from these protocols (originally presented in section 2.2) are outlined and justified here.

3.2.1 Participants

56 participants aged between 18 and 30 were recruited through the University of York Psychology electronic experiment booking system (PEEBS). All were healthy, non-smoking native English speakers with no prior history of sleep or psychiatric disorders. All participants agreed to abstain from caffeine and alcohol over the course of the experiment and 24 hours beforehand. Written informed consent for participation was obtained in compliance with the local ethical committee.

Eight participants were removed from analysis due to: (i) failure to complete training (n=5), (ii) excessively fast RTs at test (n=1, >2.5 standard deviations from the group mean), and; (iii) excessively slow RTs at test (n=2, >2.5 standard deviations from the group mean). The remaining participants were assigned to three conditions: Immediate (n=16, twelve females, mean age = 20.9, *SD*=2.8), Sleep (n=16, four females, mean age = 20.5, *SD*=3.1), or Wake (n=15, eight females, mean age = 18.9, *SD*=1). Participants had an overall mean age of 20 years (*SD*=2.55) and were provided payment or course credit for taking part, with an additional £20 Amazon gift voucher awarded to the best overall performer.

3.2.2 Design

The present experiment features a mixed-design including the between-participant variable of Delay type (Immediate, Sleep or Wake) and several within-participant variables relating to integration such as Premise type (End or Central), Inference type (non-inference, novel inference, real word inference and two unit inferences) and Size distance (Smaller or Larger distance size judgements). Participants were randomized into sleep and wake groups following a standard 12:12 sleep/wake design (Clemens et al., 2005; Tamminen et al., 2010; Wilhelm et al., 2011) while the third group of participants tested Immediately (after 20 minutes) were added as a later cohort.

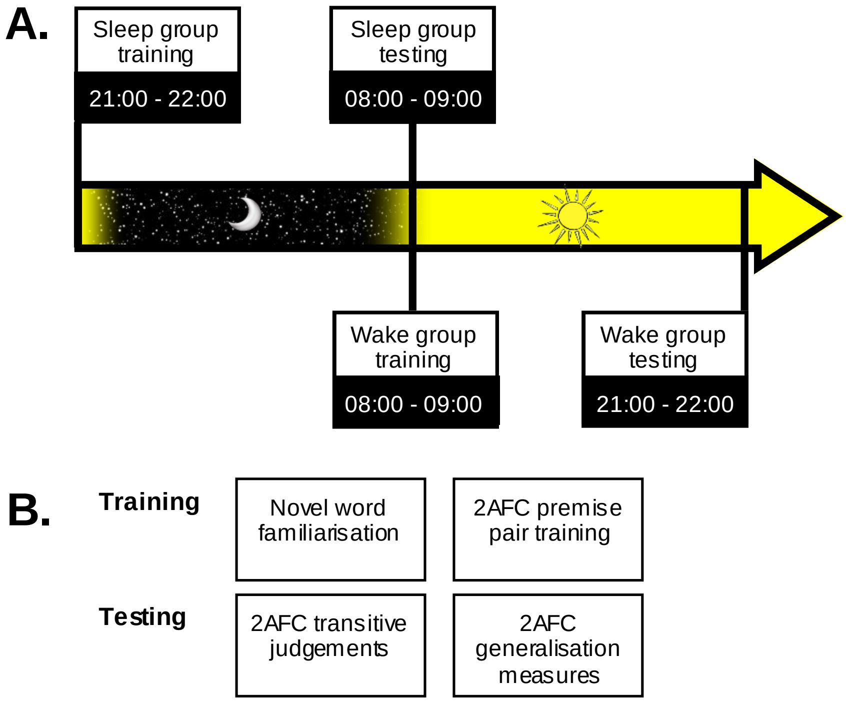


Figure 17. Experiment 2 design. (A). A 12:12 sleep/wake design. A third immediate condition was also tested 20 minutes after training. (B) The sequential order of training and testing tasks. 2AFC refers to the two alternative choice size judgments (see text).

Participants in the Wake condition started the training phase between 9:00 am and 10:00 am in the morning. After completion they spent the rest of the waking day performing normal daily activities, and returned to the laboratory between 9:00 pm and 10:00 pm in the evening to carry out the testing phase. Conversely, participants in the Sleep condition arrived at the laboratory in the evening after dinner, between 9:00 pm and 10:00 pm and underwent preparation for polysomnography (PSG) recordings prior to the training phase. Once training was completed, participants were immediately put to bed between 11 and 12 pm and woken up by the experimenter at around 7:45 am (after ~8hrs sleep) with the option of a light snack and a shower. Morning testing occurred between 30 and 40 minutes after arousal to avoid the non-specific cognitive detriments of sleep inertia (Gais et al., 2002; Tamminen, 2010). The training and testing of participants in the immediate condition, as with Experiment 1, occurred throughout a normal working day (9am - 5pm).

3.2.3 Stimuli

Overview

The most prominent alteration to the paradigm established in Experiment 1 relates to the removal of the incongruent condition so that both thematic hierarchies (animals and objects) were trained in a congruent manner, as seen in Figure 18. Additional changes to the paradigm were made to address several methodological issues raised during analysis of the previous experiment.

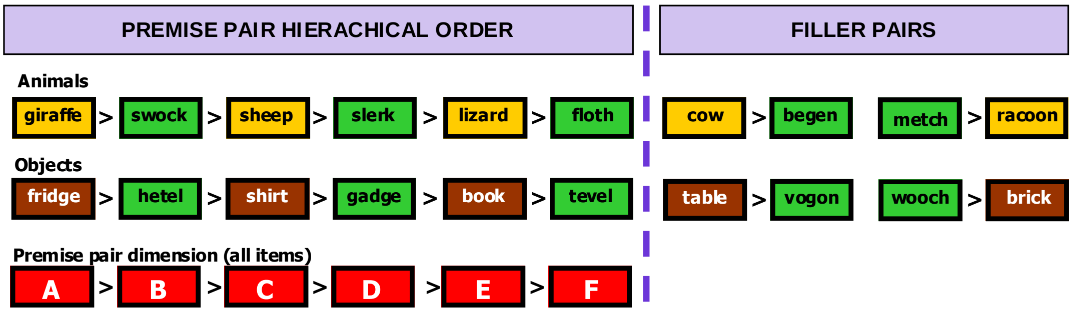


Figure 18. Experiment 2: Training hierarchies. Each participant learned animal or object premises following a six element schematic hierachy (red boxes). All green boxes represent novel word items, while the yellow and brown boxes represent animals and objects respectively. The order of presentation (animals or objects trained first) was counterbalanced across training.

Word items

All of the novel and real word items used in Experiment 1 were retained in the present experiment with the exception of the novel word item ‘creal’ (bigram frequency = 12169) which was deemed overly similar to the real word item ‘creel’ (a type of basket). As a precaution this novel word item was replaced with an equivalent novel word ‘slerk’ (bigram frequency = 12620) generated using the software WordGen (Duyck et al., 2004)(Figure 18, Animals).

3.2.4 Procedure

All instructions and experimental tasks were presented in a quiet testing room on a 17.1 inch computer monitor using E-prime software (Psychology Software tools, Inc., Pittsburgh, PA) and followed the ‘explorer scenario’ outlined in Appendix 3. The format and presentation of training and testing were identical to the procedures in used in Experiment 1 excluding the two exceptions outlined below.

Transitive inference testing

The number of blocks in the primary testing phase was expanded from four (two animal, two object) to ten in the present investigation (five animal, five object). This was done to increase the number of participant scores contributing to the measures of interest (central premises, inferences and end pair comparisons). As with Experiment 1, these additional blocks were interleaved with each other until completion, with the primary testing phase now lasting ~25 minutes.

Size distance testing

Three alterations were made to the size distance measure. Firstly a third series of ‘equivalent’ size comparisons were introduced (Table 7, Comparison 3). These items were derived from the original rated size bands (Appendix 1) and were introduced to increase the number of comparative scores contributing to this measure (as above). Secondly, untrained real words were compared to other untrained real words (real word vs. real word comparisons) to provide a baseline measure for the potential expression of a semantic distance effect. These were intermixed with an equal number of novel (novel word vs. real word) size judgments. Thirdly, comparisons in this phase were allocated ‘right’ and ‘wrong’ answers (unbeknownst to the participants) with ‘correctness’ derived from the size bands used to form the original categories. It should be noted, however, that because some of these stimuli naturally vary in size more than others (e.g. ‘dog’) accuracy measures should be considered provisional, not absolute.

Table 7. Experiment 2: Trained items and their size equivalents used in the generalisation phase.

Theme	Rank/ Size band	Trained items	Comparisons 1	Comparisons 2	Comparisons 3•
Animal	A	giraffe	bear	camel	ox
	B	swock	zebra	lion	tiger
	C	sheep	dog	pig	goat
	D	slerk	iguana	cat	monkey
	E	lizard	sparrow	squirrel	rabbit
	F	floth	worm	flea	ant
Object	A	fridge	car	cabinet	bathtub
	B	hetel	door	bike	toilet
	C	shirt	dustbin	radio	clock
	D	gadge	doll	jug	vase
	E	book	notebook	slipper	sandal
	F	tevel	envelope	glove	lightbulb

Note: the symbol '•' denotes the new comparisons introduced in this experiment

Each novel item was compared once to all real word items from comparison lists 1, 2 and 3, excluding items within the same size band (e.g. 'swock' would not be compared to zebra, lion or tiger which were approximately rated as the same size). The removal of within size band judgments was done to restrict the overall amount of ambiguous comparisons, thus allowing for a more definitive coding of 'correct' or 'incorrect' responses.

The six novel words were compared to fifteen thematically similar real words counterbalanced to avoid a screen location bias (90 comparisons total). Similarly, each real word item was compared to every other member of its comparison group resulting in 45 real word trails (90 when counterbalanced). The number of trials randomly dispersed across the two generalisation blocks therefore increased from 144 (as in Experiment 1) to 180 (~18 minutes).

3.2.5 PSG recordings and analysis

PSG recordings were obtained using a Grass Technologies system with impedance levels, filter settings and sampling rates calibrated in accordance with the standardized American Academy of Sleep Science Manual (AASM, 2007). Recordings were made digitally using an Embla© N7000 headbox and reusable Biosense© brand silver-disk electrodes. Electrode placement followed the standardized 10-20 scheme (Jasper, 1958) using a six channel montage with four bilateral scalp electrodes (O1, O2, C3, C4) each referenced to the contralateral mastoid (M1, M2). This montage was selected on the basis on both its presence in the published literature (Mednick et al., 2013; Payne et al., 2009; Wilhelm et al., 2011) and its relative simplicity, with electrode placement along the Cz midline reportedly providing a good measure of spindle activity (Cox et al., 2012; Gais et al., 2002; Schabus et al., 2004). Two electro-oculographic (EOG) channels were used to monitor eye movements (LOC, ROC), and two electromyographic (EMG) channels monitored chin movements (X1 and X3). Montage sites were cleaned using NuPrep© exfoliating gel and electrodes were affixed by means of EC2© electrode cream and surgical tape. All electrodes had impedance levels less than 5kOhms with the recording signal digitally sampled at 200Hz.

PSG Scoring

Two records were removed from analysis due to bilateral electrode removal during the night (n=1) and excessive sweat artefacts rendering the sleep record too noisy to score (n=1). Sleep scoring on the remaining 14 records was conducted manually using Remlogic© 1.1 software using the referenced C4-M1 channel in 30 second epochs. Scoring adhered to the guidelines established by Rechtschaffen and Kales (1968) with stages 3 and 4 NREM sleep, for the purposes of this investigation, collapsed into a single SWS stage. Subsequent analyses are therefore restricted to REM sleep, stage 2 sleep, and SWS.

It should be noted that a subset of three randomly records were blind coded using the R&K system by a second researcher. Average agreement levels were at an acceptable level of ~80% (Silber et al., 2007).

Sleep spindle detection

Sleep spindles were detected using an algorithm originally developed by Ferrarelli et al. (2007) which has been successfully implemented in a number of published studies (Nishida & Walker, 2007; Tamminen, 2010). EEG recordings were artefact rejected before being bandpass filtered (12-15 Hz) using a linear finite impulse response filter in EEG lab toolbox for Matlab (http://www.sccn.ucsd.edu/eeglab/). Spindle data was extracted using Matlab (The MathWorks Inc, Natick, MA) based on the number of discrete spindle events for each channel. Amplitude fluctuations in the unfiltered time series exceeding a specific threshold were counted as spindles (Figure 19). This threshold was established as eight times the amplitude in the signal of a given EEG channel, so as to account for individual variations in spindle expression (Ferrarelli et al., 2007).

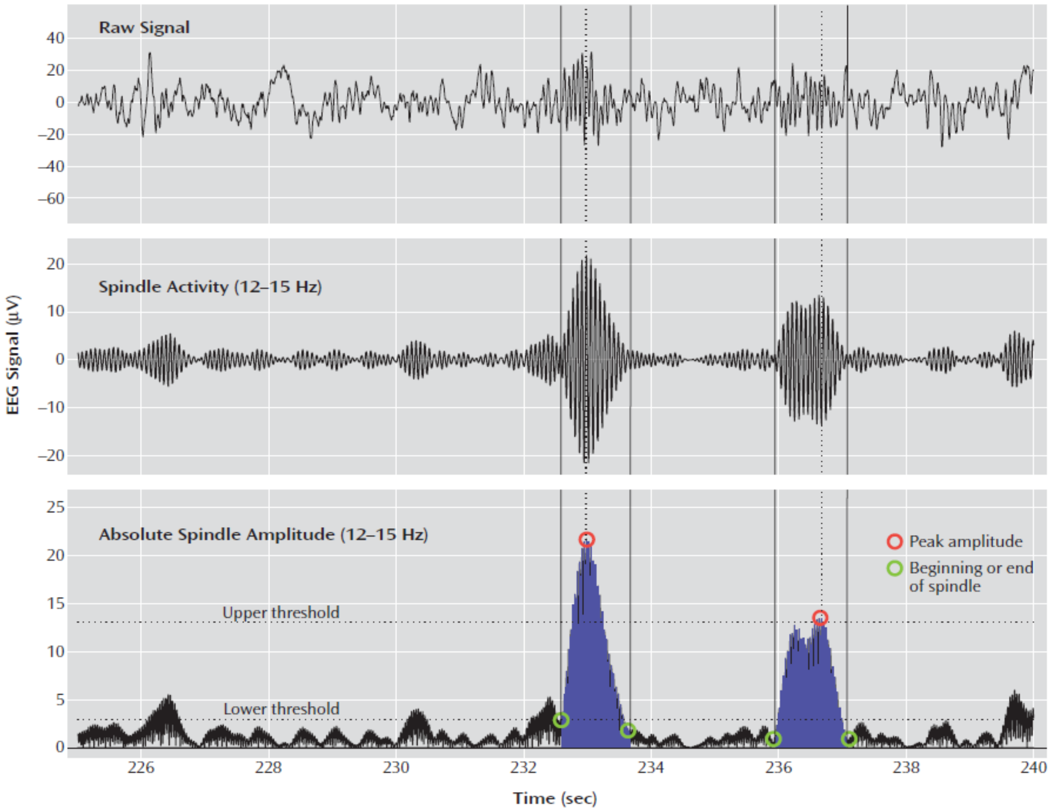


Figure 19. Experiment 2: Method for sleep spindle detection. Green and red circles indicate detected spindles. Vertical black lines enclose detected spindles in the 12-15 Hz range. This figure has been taken from Ferrarelli et al. (2007).

3.2.6 Other measures

To monitor the potential influence of time of day effects Stanford Sleepiness Scale ratings were obtained before training and testing to assess subjective levels of alertness over the course of the experiment (Hoddes et al., 1973)(Appendix 5). Average self-rated levels of sleep were recorded for the night prior to training and testing (the immediate and wake conditions) or the night prior to testing (the delayed condition only), as well as the levels of sleep normally obtained in the two weeks prior to the experiment. These were supplemented by an analogue scale (rated 1-10 from worst to best) regarding the quality of the night of sleep prior to testing for all three delay types (Appendix 4).

3.3 Results

Analysis was conducted using SPSS version 18.0, with a two-tailed $p<.05$ threshold for significance. All reaction times (RTs) are reported in milliseconds (msec) and, as with Experiment 1, excessively fast (<300 msec) or slow (>5000 msec) RTs were trimmed from the data prior to analysis (~1%).

3.3.1 Group equivalency

To ensure equivalency between the three experimental conditions comparative analyses between the three time delays was undertaken with regards to the training performance, sleep patterns and alertness levels throughout the experiment.

Training

The number of blocks required to complete training (between 4-10) and the average accuracy levels on these final blocks (≥80%) were independently analyzed using a one-way ANOVA featuring the between-participants variable of Delay type (Immediate vs. Wake vs. Sleep). These analyses indicated that the number of completed blocks required to exit training were not significantly different between the Immediate (M=5.3, SE=.32), Wake (M=5.9, SE=.34) or Sleep (M=5.2, SE=.19) conditions, F(2,44)=1.43, p=.249. Furthermore, accuracy levels on these final blocks remained equivalent when performance was aggregated across all premise pair items (M=91%, SE=.59), F<1, and central premises alone (M=83%, SE=1.39), F<1, thus indicating equivalent training performance across all three time delays (Table 8).

Table 8. Experiment 2: Training performance.

	Immediate group		Wake group		Sleep Group	
	All premises	Central premises	All premises	Central premises	All premises	Central premises
Accuracy (%)	91 (1)	83 (2.3)	91 (1.3)	85 (2.8)	90 (1)	83 (2.3)

Note: 'All premises' includes filler items. Central premises refers to performance on B>C, C>D and D>E items only. Values in parentheses denote the standard error.

Sleep measures

Participants provided self-rated measures as to (i) the average amount of sleep obtained two weeks prior the experiment, and (ii) the amount of sleep acquired the night before testing. The dependent variable of sleep amount (in hours) was entered into a mixed-effects ANOVA featuring the between-participants variable of Delay type (Immediate vs. Wake vs. Sleep) and the within-participants variable of Sleep period (Amount of sleep obtained normally vs. Amount obtained on the experimental night).

These data, which can be found in Table 9, indicated no main effect of Delay type, F(1, 44)=2.31, p=.111, nor a main effect of Sleep period, F(1, 44)=1.15, p=.289, thus highlighting consistent sleep patterns across the three Delay types, both in terms of regular sleep patterns (M=8.13 hours, SE=.12), and the amount of sleep obtained on the experimental night (M=7.98 hours, SE=.13). Furthermore, a Delay type x Sleep period interaction was not present, F(2,44)=1.05, p=.360, indicating that the sleep acquired on the experimental night did not diverge significantly from regular sleep patterns for any of the three delay types.

Table 9. Experiment 2: Self-reported sleep patterns.

Immediate group		Wake group		Sleep Group	
Usual sleep	Sleep on experimental night	Usual sleep	Sleep on experimental night	Usual sleep	Sleep on experimental night
8 (.21)	8 (.22)	8 (.22)	7.7 (.22)	8.3 (.2)	8.4 (.22)

Note: The amount of sleep is reported in hours. Values in parentheses indicate the standard error.

An additional one-way ANOVA was conducted between the three Delay types (Immediate vs. Wake vs. Sleep) on the perceived quality of the sleep obtained on the experimental night (i.e. prior to testing) on a 1-10 scale (from worst to best). Again, no differences emerged on this measure between the Immediate (*M*=6.7, *SE*=.5), Wake (*M*=7.3, *SE*=.5) and Sleep groups (*M*=6.3, *SE*=1.9), *F*<1. This result is in accord with the above analysis regarding sleep patterns, indicating that all three Delay types acquired nights of sleep with comparable quality before testing.

Alertness

Lastly, alertness levels were assessed regarding the self-reported measures of sleepiness on the Stanford Sleepiness Scale (SSS) which were obtained prior to the training and testing phases. A 3 x 2 mixed-effects ANOVA was conducted featuring the between-participants variable of Delay type (Immediate vs. Wake vs. Sleep) and the within-participants variables of Phase (Training vs. Testing) using the dependent variable of SSS scores. No main effect of Delay type, *F*<1, or Phase, *F*(1,44)=1.86, *p*=.180, was apparent. However, a marginally significant interaction between these factors emerged *F*(2,44)=2.78, *p*=.073. Cursory examination of the SSS ratings in Table 10 indicate that this interaction was driven by slightly lower ratings of alertness at the start of training in the Sleep condition relative to the other two Delay types, most likely a result of fatigue due to the later time of training (~09:30 pm). Nevertheless, average SSS ratings (*M*=2.4, *SE*=.13) indicated equivalent levels of moderate alertness throughout the experiment across all three Delay types (Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973).

Table 10. Experiment 2: Stanford Sleepiness measures.

Phase	Immediate group	Wake group	Sleep Group
Training	2.1 (.27)	2.5 (.27)	2.9 (.27)
Testing	2.3 (.28)	2.4 (.29)	2.3 (.28)

Note: Values in parentheses indicate the standard error.

3.3.2 Premise pair performance and the serial position effect

The aim of these analyses was to assess memory integration in relation to the serial position effect typically expressed through weaker memory performance for items central to the schematic hierarchy relative to end items (Greene et al., 2001; Moses et al., 2006; Ryan et al., 2009; Wynne, 1995). Following on from the observations made in Experiment 1, central premises are predicted to be enhanced after a period of sleep counter to the direction of the serial position curve associated with this effect. As with Experiment 1, premises pairs were aggregated into measures of Central (B>C, C>D, D>E) and End premise performance (A>B, E>F) to form the factor of Premise type. Accuracy and accurate RT measures for this factor were then entered into 3 x 2 mixed-effects ANOVAs featuring the between-participant variable of Delay type (Immediate vs. Wake vs. Sleep) and within-participant variable of Premise type (End vs. Central).

Accuracy levels and the serial position effect

Analysis of the accuracy levels across premises, which are displayed in the top half of Figure 20, indicated no main effect of Delay type, *F*<1, thus illustrating equivalent accuracy levels across the three time delays. A large main effect of Premise type was present, with higher levels of accuracy for End premises (*M*=88%, *SE*=1.43) relative to Central premises (*M*=68%, *SE*=2.38), *F*(1,44)=46.59, *p*<.001, indicating a pattern of performance in line with the serial position effect (Wynne, 1995). However, this finding was not qualified with a Delay type x Premise type interaction, *F*<1. As such, despite the presence of a serial position effect, there was little to distinguish the three Delay types regarding its expression.

Accurate RT performance and the serial position effect

Accurate RT levels, which are displayed in the bottom half of Figure 20, similarly did not reveal a main effect of Delay type, *F*<1, indicating equivalent RT performance across the three time delays. As with the analysis of accuracy levels above, a large main effect of Premise type was apparent, *F*(1,44)=126.21, *p*<.001, driven by faster responses for End (*M*=1078, *SE*=40.32) relative to Central premises (*M*=1440, *SE*=52.47), providing additional evidence for the presence of a serial position effect. However, this main effect was not qualified with a Delay type x Premise type interaction, *F*(2,44)=1.78, *p*=.180, indicating that, counter to predictions, the serial position effect remained undifferentiated by Delay type.

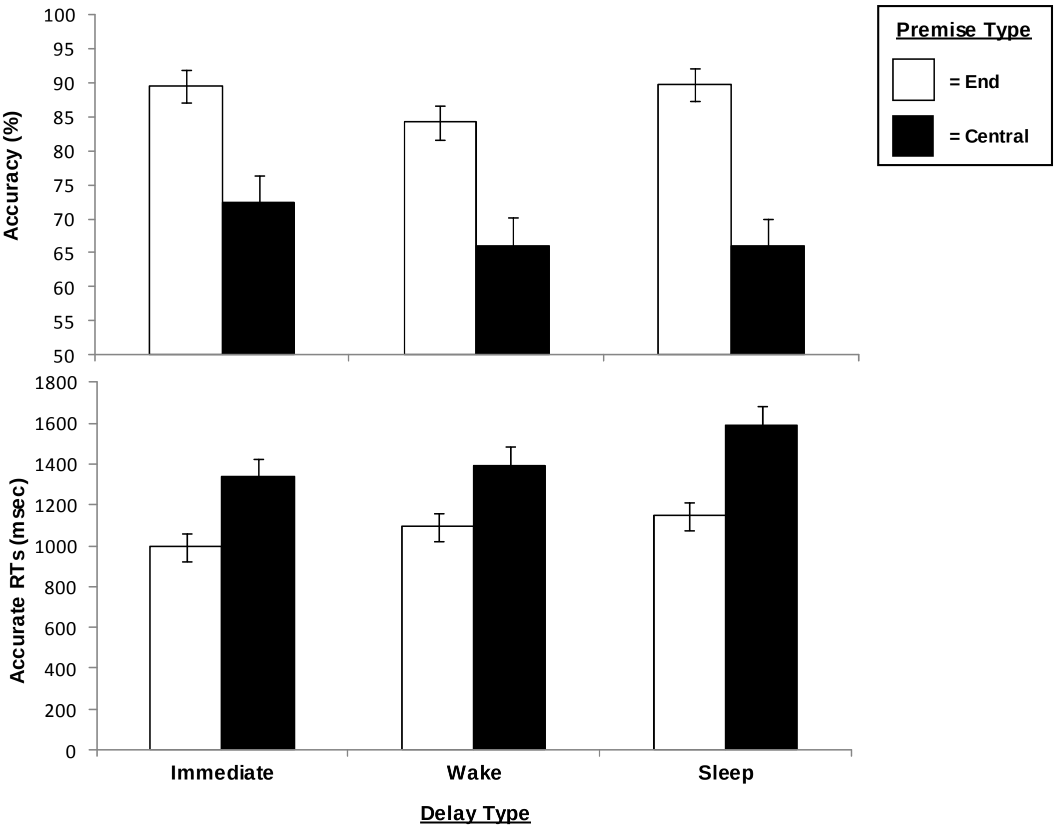


Figure 20. Experiment 2: Accuracy and accurate reaction times for end and central items across all three experimental groups. Error bars represent the Standard Error.

3.3.3 Transitive inferences and memory abstraction

As with Experiment 1, TI analysis targeted four comparisons of interest: the non-inference control comparisons (A>F), the one unit novel item transitive inferences (B>D), the one unit real item transitive inferences (C>E), and the two unit transitive inferences (B>E), which together formed the factor of Inference type. This factor was entered into a 3 x 4 mixed-effects ANOVA using the between-participants variable of Delay type (Immediate vs. Wake vs. Sleep) and within-participants variable of Inference type (A>F, B>D, C>E and B>E) as measured through accuracy (%) and accurate RT performance. Mauchly's test indicated that the assumptions of sphericity were violated across Inference types for both accuracy, $\chi_2(5) = 31$, *p*= <.001, ε=.656, and accurate RT measures, $\chi_2(5) = 20.42$, *p*= .001, ε=.822. Results are therefore reported using a Greenhouse-Geisser correction where appropriate.

Transitive inferences and accuracy levels

Accuracy levels on this measure, which are illustrated in Figure 21, revealed a large main effect of Inference type, *F*(1.97, 86.55)=23.33, *p*<.001, however as this difference was expected, and is not of interest to the present analysis, it will not be explored further. No main effect of Delay type was present, *F*<1, nor was there an interaction between Delay type and Inference type, *F*<1, leaving little to distinguish the three time delays in terms of accuracy performance.

Transitive inferences and accurate RT performance

Analysis of accurate RTs, which are also illustrated in Figure 21, revealed main effect of Delay type with a borderline level of significance, *F*(2, 42)=3.23, *p*=.05. Post hoc bonferroni-corrected comparisons between the three Delay types indicated the Sleep condition was marginally slower than the Immediate condition (*Mean difference* = 242 msec, *p*=.056) whereas no other comparisons differed significantly, *P*s>.225. Visual inspection of Figure 21 indicates that this difference was most prominently expressed in the relative slowing in the RTs for Real word inferences and two unit TIs; whereas RTs for the critical novel word inferences remained equivalent. A large main effect of Inference type was present, *F*(2.53, 106.32)=54.65, *p*<.001, however this finding was not qualified with an interaction between Delay type and Inference type, *F*(6,126)=1.31, *p*=.257. Taken together, these results do not support the hypothesis that a period of sleep will result in the selective enhancement of novel word TIs. Indeed the general slowing of responses, particularly in relation to real word TIs, suggests that some kind of circadian confound may be present in the sleep condition (Monk & Leng, 1982; Tilley & Warren, 1983).

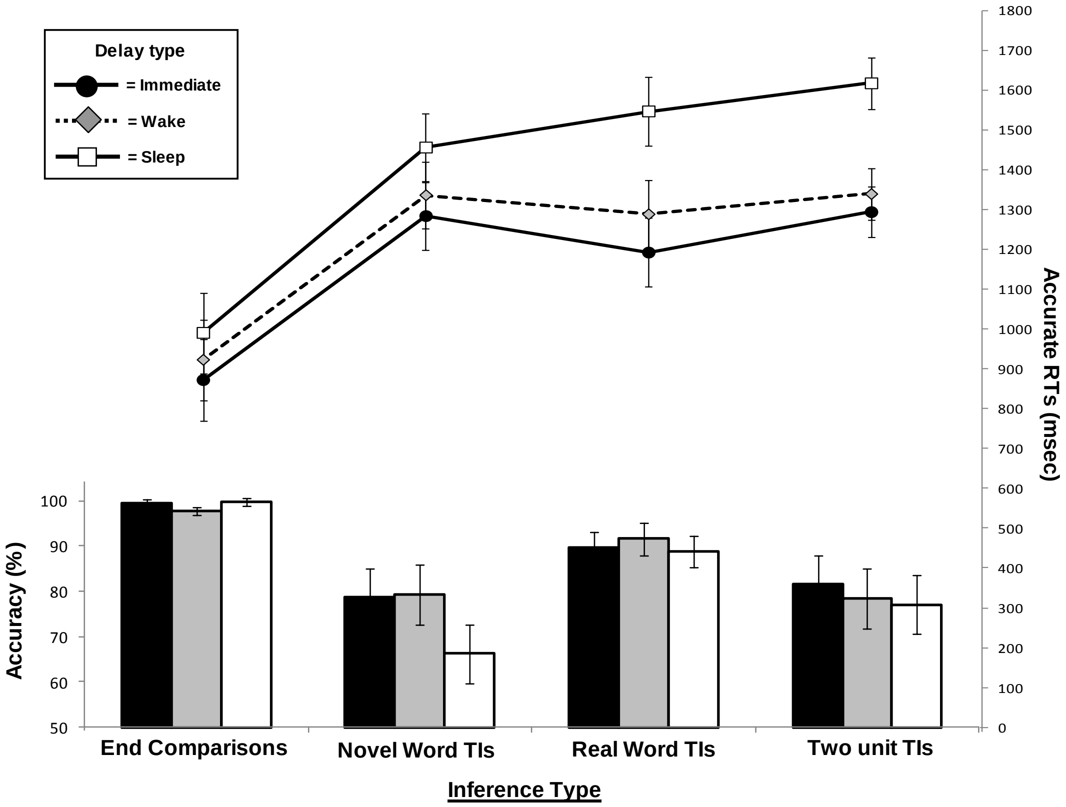


Figure 21. Experiment 2: Transitive Inferences across all testing blocks. End comparisons = A>F, Novel Word TIs = B>D, Real Word TIs = B>D, and Two unit TIs = B>E. Error bars represent the standard error.

3.3.4 Size generalisations and the semantic distance effect

Recall from Chapter 2 that the generalisation measure relates to the expression of the semantic distance effect. This effect is typically expressed in the faster (and potentially more accurate) responses for size judgements made between items that are relatively distant in size (e.g. bear>zebra) compared to size judgements between items that are closer together in size (e.g. bear>worm). The generalisation measure in the present experiment features several improvements compared to the equivalent measure in Experiment 1. Changes include the introduction of accuracy levels accompanying RT performance and, importantly, a baseline measure of the semantic distance effect in terms of size judgements exclusively between real word items. With regards to consolidation, the expression of the semantic distance effect is predicted to be more pronounced for generalisations involving novel word items after a period of sleep, whereas size judgements occurring exclusively between real word items are predicted to remain undifferentiated across the three time delays. The two different types of size judgement (Novel word and Real word judgements vs. Real word only judgements) form the factor of Judgement type in the present analyses.

As with Experiment 1, size distances were aggregated into 'Smaller' and 'Larger' designations based on the differences between size band allocations (ranging from 1 to 5). Comparisons for the two closest distances formed the Smaller proximity (1 and 2) and the remaining three size distances (3-5) formed the Larger proximity, ensuring an equal number of size judgments scores between the Smaller and Larger proximity measures. Analyses implemented a 3 x 2 x 2 mixed-effects ANOVA featuring the between-participant variable of Delay type (Immediate vs. Wake vs. Sleep) and the within-participant variables of Judgement type (Novel word and Real word judgements vs. Real word only judgements) and Proximity (Smaller vs. Larger) using the dependent variables of accuracy (%) and accurate RTs (msec) for these measures.

Accuracy levels and the semantic distance effect

The analysis of accuracy performance, which can be found in Table 11, did not reveal a main effect of Delay type, *F*<1, indicating that the levels of generalisation accuracy were equivalent across the three time delays. A large main effect of Judgement type was expressed across all conditions, *F*(1, 44)=197.38, *p*=.001, driven by higher levels of accuracy for size judgments exclusively between real words (*M*=92%, *SE*=.73) relative to those between novel words and real words (*M*=76%, *SE*=1.19). An additional main effect was found for Proximity, *F*(1, 44)=86.84, *p*=.001, indicating higher levels of accuracy for comparisons involving Larger distances (*M*=87%, *SE*=.75) relative to Smaller distances (*M*=81%, *SE*=.98), strongly suggesting the presence of a semantic distance effect for both Judgement types (Moyer, 1973; Pavio, 1975).

However, no interaction was present between Proximity type and Judgment type, *F*<1, indicating that the expression of the semantic distance effect remained undifferentiated by this factor. Moreover, in terms of the experimental predictions, no interactions were present between Delay type and Proximity, *F*<1, Delay type and Judgement type, *F*<1, or three way interaction between these variables, *F*<1. As such, despite the presence of a semantic distance effect for both Judgment types, no differences emerged with regards to the magnitude of the semantic distance effect across the three time delays.

Table 11. Experiment 2: Size generalisation measures.

		Immediate		Wake		Sleep	
Judgement type	Proximity	Accuracy (%)	RT (msec)	Accuracy (%)	RT (msec)	Accuracy (%)	RT (msec)
Novel word vs. Real word	Smaller	74 (2.3)	1441 (95)	72 (2.4)	1474 (98)	73 (2.3)	1713 (95)
	Larger	79 (2.2)	1198 (73)	78 (2.3)	1277 (76)	80 (2.2)	1484 (73)
Real word vs. Real word	Smaller	87 (1.7)	1289 (76)	89 (1.7)	1345 (79)	90 (1.7)	1613 (76)
	Larger	94 (1)	1102 (64)	96 (1)	1176 (66)	96 (1)	1342 (64)

Note: Values in parentheses represent the standard error. RT refers to accurate reaction times.

Accurate RTs and the semantic distance effect

Accurate RT analysis, which can also be seen in Table 11, revealed three significant main effects. Firstly, a main effect of Delay type was present, *F*(2, 44)=4.44, *p*=.017. Bonferroni-corrected post hoc comparisons between the three time Delays indicated that RTs were slower in the Sleep condition relative to the Immediate condition (*Mean difference* = 280 msec, *SE*=98.81, *p*=.021), with the Sleep condition also expressing a non-significant trend towards slower responses relative to the Wake condition (*Mean difference* = 220 msec, *SE*=100.44, *p*=.101) whereas no differences emerged between the Immediate and Wake conditions (*Mean difference* = 60 msec, *SE*=100.44, *p*=1). Secondly, a main effect of Judgment type was apparent, *F*(1,44)=15.39, *p*<.001. In accord with the corresponding analysis of accuracy levels above, this result indicates faster responses for size judgements exclusively between real words (*M*=1311 msec, *SE*=40.02) relative to novel word vs. real word judgments (*M*=1431 msec, *SE*=46.84). Thirdly, a large main effect of Proximity was evident, *F*(1,44)=1.11, *p*<.001. Again, as with the derived accuracy measures, RTs for both comparisons types were in the direction of a semantic distance effect with faster responses to Larger (*M*=1263, *SE*=37.34), as opposed to Smaller size distance comparisons (*M*=1479, *SE*=46.28).

Although cursory inspection of the RT measures in Table 11 suggest that novel word judgements were prominently slower for Smaller novel word vs. real word comparisons in the Sleep condition, this difference did not prove strong enough to induce an interaction between Delay type and Judgment type, *F*<1, Delay type and Proximity, *F*<1, or three-way interaction between these variables, *F*<1.

Taken together these generalisations measures indicate a semantic distance effect for both Judgement types, although responses for novel vs. real word comparisons tended to be slower and less accurate than comparisons between real words. Counter to predictions, the only factor distinguishing the sleep condition on these measures was a tendency for slower responses, a pattern of performance that was importantly found for both judgement types. The most parsimonious interpretation for this finding (discussed more in section 3.3.5), particularly in light of the slowing of the real word TIs reported in the previous section, is that some kind of circadian confound may be present in the sleep condition.

3.3.5 Sleep stages and integration

Data from the intact sleep records (n=14), as presented in Table 12, indicate overall sleep parameters within the range expected from normal healthy adults (Spriggs, 2008). The measures of stage 2, REM sleep and SWS, and total sleep time (TST) present in Table 12 were correlated with several factors derived from the three measures of integration in the previous sections, as outlined below.

Table 12. Experiment 2: sleep parameters.

Sleep parameter	Mean time (min) ± SEM	% of total sleep time ± SEM
Total sleep time	523 ± 15	
Wake after sleep onset	32 ± 16	
Sleep latency	32 ± 16	
Stage 1	52 ± 16	10 ± 1.2
Stage 2	250 ± 17	48 ± 1.3
SWS (stages 3 + 4)	79 ± 15	15 ± 1.3
REM	78 ± 15	15 ± 1.3

Note: SWS = slow wave sleep, REM = rapid eye movement sleep, SEM = standard error of the mean

Behavioural factors and multiple comparisons

Before undertaking sleep analysis in this experiment it is important to outline how the issue of multiple comparisons will be address. The exploratory nature of the investigations contained in this thesis necessitates multiple correlations between sleep physiology and a rich set of behavioural data. An inherent problem with this approach is that it brings with it an increased risk of generating spurious relationships through multiple comparisons, thus inflating the risk of a type I statistical error. Moreover, there is presently no standardised approach regarding this issue in the sleep literature, particularly for investigations containing a large number of behavioural factors.

As such two rules regarding multiple comparisons outlined here will be consistently applied to correlational analysis throughout this thesis. Firstly, the overall number of multiple comparisons required in a given experiment will be reduced by aggregating the behavioural measures down into the simplest measures possible representing the principle effects of interest. Secondly, bonferroni adjustments for multiple comparisons will be inflated by the number of behavioural factors featured in a given correlational analysis but not by the number of sleep measures. This second rule is intended to strike a balance between the risk of type I and type II statistical errors in order to preserve the integrity of any derived conclusions.

Sleep stage correlations

Following the rules outlined above performance on the central items (B>C, C>D, D>E) was subtracted the end items (A>B, B>C), to produce a single value representing the extent of the serial position effect. This factor was calculated for both accuracy (%) and RTs (msec) (see section 3.3.2). It is important to note, regarding subsequent interpretation, that for % measures a greater reduction in the serial position effect (i.e. potential integration) will manifest in more positive values on this factor, whereas a reduction in the serial position effect for RTs will produce more negative values.

A similar factor was created in relation to the expression of the semantic distance effect. Namely, performance for larger distance comparisons were deducted from smaller distance comparisons using the data featured in section 3.3.4. In terms of interpretation, positive % values indicate a greater the expression of the semantic distance effect, and thus potential integration. Conversely, for RTs, a larger semantic distance effect will be reflected in negative values on this composite measure.

Lastly, for the inference items three of the four variables featured in the primary analyses in section 3.3.3 were implemented: novel word TIs (B>D), real word TIs (C>E), and the two unit TIs (D>E). The non-inference comparison (A>F) was not included in this analysis as the asymptotic levels of performance for these comparisons did not provide enough variability for subsequent correlation (*M*=100%, *SE*=.36). As with the premises and generalisations above, both accuracy and RTs for TI performance were correlated with the sleep measures.

Correlations were established using Pearson's correlation coefficients, the products of which can be found in Table 13. As 12 behavioural factors were featured in this analysis the alpha value adjustment for the subsequent bonferroni corrections was a multiplication of 12. The p-values for these analyses are presented in Table 13 in their uncorrected form, however results that exceeded the .05 alpha threshold after correction are marked with the symbol '†'.

Table 13. Experiment 2: Correlations between integration measures and time spent in different stages of sleep.

		Sleep Stage			
Premises		TST	Stage 2	REM	SWS
Acc (%)	*r*	-0.122	-0.218	-0.175	0.281
	p	*0.68*	*0.45*	*0.55*	*0.33*
RT (msec)	*r*	-0.082	-0.154	-0.496	-0.053
	p	*0.78*	*0.60*	*0.07*	*0.86*
Transitive Inferences					
Novel word inference (%)	*r*	-0.232	-0.128	-0.239	0.337
	p	*0.42*	*0.66*	*0.41*	*0.24*
Real word inference (%)	*r*	-0.082	-0.213	-0.288	0.355
	p	*0.78*	*0.46*	*0.32*	*0.21*
Two unit inference (%)	*r*	-0.023	0.090	0.256	**0.585**
	p	*0.94*	*0.76*	*0.38*	***0.028†***
Non inference (msec)	*r*	-0.091	-0.299	-0.101	0.024
	p	*0.76*	*0.30*	*0.73*	*0.94*
Real word inference (msec)	*r*	-0.122	-0.512	**-0.552**	-0.028
	p	*0.68*	*0.06*	***0.041†***	*0.92*
Two unit inference (msec)	*r*	-0.192	-0.488	**-0.564**	-0.357
	p	*0.51*	*0.08*	***0.036†***	*0.21*
Generalisation					
RW Size difference (%)	*r*	**-0.547**	-0.031	-0.128	-0.048
	p	***0.043†***	*0.92*	*0.66*	*0.87*
NW Size difference (%)	*r*	0.006	-0.100	0.204	-0.130
	p	*0.98*	*0.73*	*0.48*	*0.66*
RW Size difference (msec)	*r*	-0.128	0.449	-0.155	-0.284
	p	*0.66*	*0.11*	*0.60*	*0.32*
NW Size difference (msec)	*r*	-0.223	0.156	0.109	**-0.586**
	p	*0.44*	*0.59*	*0.71*	***0.028†***

Note: Significant correlations in bold. TST =total sleep time, REM = time spent (in minutes) in rapid eye movement sleep, SWS = time spent (in minutes) in slow wave sleep. † = p-values that do not survive a Bonferroni correction for multiple comparisons. The unreported Bonferroni corrections inflated the alpha values by the total number of comparisons per sleep stage (12).

Correlational analysis

The two strongest results in Table 13 denote correlations with SWS. Increased SWS was firstly associated with improved accuracy for two unit inferences, and secondly, in the expression (and thus potential integration) of the semantic distance effect between size generalisations featuring novel words and real words. Although both of these effects are in line with a predicted relationship between sleep and memory performance these correlations did not survive correction for multiple comparisons, and thus should be considered tenuous.

Two weaker relationships emerged between the accurate RTs for real word and two unit TIs. In both of these cases the negative relationship indicates that increased time in this stage is associated with faster reaction times on these measures, as might be predicted by the REM sleep consolidation account introduced at the start of this chapter (Walker & Stickgold, 2010). Lastly, TST was negatively associated with the presence of a semantic distance effect in terms of accuracy counter to the direction that might be predicted with regards to memory consolidation. However, as above, these relationships did not survive the correction for multiple comparisons.

While the present analysis deals with absolute time in each sleep stage, another common way to assess the robustness of these potential relationships is to compare performance measures to the percentage of REM, SWS and stage 2 sleep as part of TST. The advantage of such an analysis is that it provides a measure of the contribution of each sleep stage independent of TST. Should the expressed association between a given sleep stage and behavioural measure represent a true relationship; we might expect it to be expressed both in terms of percentage of TST as well as the amount of time spent in that stage.

This analysis was undertaken (see Appendix 6) and the only correlation to be reproduced was the positive relationship between SWS and two unit TI accuracy, *r*(14)=.604, *p*=.022 †. While this result, too, did not survive correction for multiple comparisons, the presence of this relationship across both types of sleep measure suggests that it may be more robust. Equally, as the correlation between the RTs for novel word size generalisations and SWS was not re-expressed, this relationship is potentially more equivocal with regards to the time spent in these stages.

3.3.6 Sleep spindles and integration

As sleep spindles in both stage 2 (Gais et al., 2002; Schabus et al., 2004, 2006) and SWS (Cox et al., 2012; Wilhelm et al., 2011) have been found to contribute independently to declarative memory consolidation, analysis was conducted on each of these stages separately. Furthermore, reported analyses are restricted to measures of spindle density (spindles per minute) as this variable provides a measure of spindle activity controlling for time. The high correlation between spindle density and absolute spindle count, *r*(14)=.831, *p*<.001, indicates that these variables are closely affiliated, and are essentially the same measure (Tamminen, 2010).

Overview

As an overview, spindle density measures (both count and density) are presented in Table 8 in relation to electrode placement. It should be noted that levels of spindle activity in this figure, both in terms of count and density, are generally lower that those reported in other word learning investigations, including studies that have implemented the same spindle detection algorithm (cf. Clemens et al., 2005; Schabus et al., 2004; Tamminen et al., 2010). The reasons for this will be discussed more fully in section 3.4.1, but most probably relate to the over-priming of semantically congruent information during training in this experiment, a factor that appears to have diminished the potential influence of consolidation.

Table 14. Experiment 2: Sleep spindle measures across sleep stage and electrode site.

Sleep stage	Electrode site	Total N	Density
Stage 2	C3	181 (22)	0.76 (.08)
	C4	154 (25)	0.63 (.07)
SWS	C3	22 (5)	0.27 (.05)
	C4	24 (5)	0.27 (.06)

Note: Standard error in parentheses. Total N = total number of spindles averaged across participants, Density = spindle density (number of spindles per minute). SWS = slow wave sleep.

Behavioural correlations

Spindle density was correlated with the same behavioural measures of integration assessed in the primary sleep stage analysis in Table 13 using a series of Pearson's correlation coefficients. Given the lack of firm statistical differences in spindle activity across the bilateral electrode sites, and to reduce the number of multiple comparisons in these analyses, spindle density across the C3 and C4 electrode sites were collapsed into composite measure. The resulting correlations can be found in Table 15 with results that exceeded the .05 alpha threshold after correction marked with the symbol '†'.

These correlations revealed three relationships of interest. Firstly, in line with the experimental predictions, SWS spindle activity was predictive of the relative integration of premise pairs. This was evidenced exclusively for the accuracy factor relating to the serial position effect, which displayed a positive relationship with SWS spindle density. This correlation provides provisional evidence for the active maintenance of new information in a manner that contradicts the expression of the serial position effect. For the TI measures, again, SWS density was positively associated with accuracy levels, although this relationship was restricted to novel word TIs only. Despite the lack of differences in the memory performance of novel word TIs between the three time delays (see section 3.3.3), this result suggest that SWS spindles may nevertheless support the expression of these inferences. Interestingly, the real word TIs also shared a positive relationship with spindle activity, albeit with spindles in stage 2 sleep only, indicating that TIs between pre-existing words may also benefit from active consolidation. No relationships were evident for any of the measures relating to the semantic distance effect.

Table 15. Experiment 2: Spindle activity in relation to integration measures.

		Spindle density	
Premises		Stage 2	SWS
Acc (%)	*r*	0.412	**0.603**
	p	0.14	**0.022**†
RT (msec)	*r*	-0.335	-0.415
	p	0.24	0.14
Transitive Inferences			
Novel word inference (%)	*r*	0.506	**0.580**
	p	0.06	**0.030**†
Real word inference (%)	*r*	**0.599**	0.52
	p	**0.024**†	0.06
Two unit inference (%)	*r*	0.386	0.479
	p	0.17	0.08
Novel word inference (msec)	*r*	0.039	0.228
	p	0.90	0.43
Real word inference (msec)	*r*	-0.399	-0.310
	p	0.16	0.28
Two unit inference (msec)	*r*	-0.226	-0.115
	p	0.44	0.70
Generalisation			
RW Size difference (%)	*r*	0.017	-0.453
	p	0.95	0.10
NW Size difference (%)	*r*	-0.156	-0.395
	p	0.60	0.16
RW Size difference (msec)	*r*	0.307	0.062
	p	0.29	0.83
NW Size difference (msec)	*r*	-0.373	-0.342
	p	0.19	0.23

Note: Significant correlations in bold. SWS = spindle activity in slow wave sleep. Spindle count = total n of spindles in a given sleep stage. Spindle density = number of spindles per minute . † = p-values that do not survive a Bonferroni correction for multiple comparisons. The unreported Bonferroni corrections inflated the alpha values by the total number of comparisons per sleep stage (12).

Although none of the three relationships revealed in this analysis survived correction for multiple comparisons it should be noted that the majority of the correlations in Table 15, even those that failed to reach any kind of significance threshold, are in a direction that would be predicted by the active consolidation model. As such, the present relationships regarding spindle activity may be genuine. Taken together the three correlations found here provide provisional support for the active maintenance of declarative memory during sleep, although this maintenance effect is not likely a unique contribution of the sleep state considering the null behavioural results reported in the previous sections.

3.4 Summary and Discussion

To explore the potential mechanisms involved in the consolidation of newly learned, yet semantically compatible information, the present experiment employed an adapted version of the TI paradigm to probe memory integration in relation to sleep physiology. Following on from the results of Experiment 1, it was predicted that integration would be more prominent after a 12 hour delay involving sleep, as opposed to a 12 hour period of wakefulness or a 20 minute delay. Furthermore, memory integration, as measured by premise pair, TI and size generalisation measures, was predicted to display a relationship with one or more aspects of sleep architecture, in particular SWS and its affiliated spindles.

Overall, the results of the present investigation to do not reveal the same pattern of integration observed in Experiment 1, and are indeed marked by a distinct lack of consolidation-associated integration as might have been registered, in particular, by RT measures relating to central premise and novel TI performance. The reason for this, which is outlined in more detail in the following section, is most likely due to the increased exposure to congruent information in the present investigation relative to Experiment 1. This factor appears to have accelerated RTs beyond a threshold sensitive to integrative consolidation.

Although there is minimal evidence for the predicted qualitative shifts in behaviour, instead there is provisional evidence for more conventional forms of quantitative consolidation, namely in the maintenance of information across the sleep period. Broadly speaking, evidence supporting this interpretation is threefold. Firstly, the majority of integrative measures, particularly when assessed through accuracy, revealed few differences between the three time delays, suggesting the minimal forgetting of learned materials over 12 hours of sleep and wakefulness. Secondly, and in accordance with the original experimental hypotheses, several relationships emerged between sleep architecture and these behavioural measures, particularly those associated with central premises and TI performance. Thirdly, the majority of these relationships, although weak, were associated with SWS spindle activity, a known biomarker of systems-level consolidation.

Consequently, although the maintenance of information may not be a unique contribution of the sleep state *per se*, aspects of the sleep architecture may nevertheless support the flexible expression of memory. However, these results are interpreted with an additional caveat regarding a possible circadian confound present in the sleep condition, which was expressed through a slowing of real word baseline comparisons in both the generalisation and TI measures.

3.4.1 The lack of integration in Experiment 2

Firstly, the failure to replicate the pattern of integration observed in Experiment 1 needs to be addressed. Specifically, after a period of sleep there was no predicted acceleration of RTs for novel word TIs or central premise items. Although Experiments 1 and 2 examined performance across different time courses, ongoing theoretical and experimental accounts of consolidation would nevertheless predict the largest degree of integration after a period of sleep. Why was this not the case?

When the results are compared between Experiments 1 and 2 it quickly becomes evident that: (i) the present experiment yielded considerably faster RTs in almost every measure, particularly those found previously to be sensitive to integration, and; (ii) these speeded responses were apparent in the first 20 minutes after training and largely persisted across both sleep and waking time delays. This point is illustrated neatly by comparing congruent novel word TIs in the immediate condition between Experiments 1 and 2. In the present investigation accurate RTs on these items were faster by a striking ~800 msec (*SED* = 187.91), $t(30)=4.26$, $p<.001$, an effect size greater than the integration benefits observed across the entire 24 hour delay in Experiment 1 (*Mean difference =* ~ 500 msec, *SED*= 244.32). These accelerated RTs cannot be attributed to the expanded number of testing blocks featured in Experiment 2 (which increased from 4 to 10), which might have otherwise provided a degree of additional training at test, as further (unreported) analysis examining the first two blocks of testing only revealed a similar pattern of results. Overall, it is therefore apparent that the behavioural markers used to probe integration were already expressed prior to a consolidation opportunity.

The reason for this most likely relates to the key difference between Experiment 1 and Experiment 2, namely that here *both* animal and objects items were trained in a semantically compatible manner (e.g. the ordering 'giraffe > swock > sheep > slerk > lizard > floth'), whereas in Experiment 1 half of these items featured conflicting semantic information (e.g. the novel word 'slerk' across the premises 'sheep > swock > lizard > slerk > giraffe > floth'). Although participants in both experiments were broadly equivalent in terms of their initial training performance, the saliency of congruent information in the present experiment may have been considerably more pronounced given that training essentially featured double the amount of congruent information.

There is now considerable evidence that consolidation processes may be particularly effective for weaker, or more difficult to encode information (Drosopoulos, Schulze, et al., 2007; Kuriyama et al., 2004) including visualized relationships between real word items (Schmidt et al., 2006) and the ability to abstract shared features across semantically linked elements of information (Diekelmann et al., 2010). The most parsimonious explanation for the lack of consolidation-associated integration in Experiment 2 is that the heightened saliency of congruent semantic information accelerated RTs beyond a threshold sensitive to such effects (Stickgold, 2009).

An additional point relates to the absence of incongruent training, which may have otherwise induced a degree of confusion regarding the ordinal relationships across novel word stimuli. The lack of this manipulation relative to Experiment 1 may have enhanced participant's explicit awareness of the overarching stimuli relations (i.e. 'A>B>C>D>E>F'). In the TI literature it has been found that when such explicit awareness develops, levels of performance on both premises and TIs improve rapidly (Greene et al., 2001; Libben & Titone, 2008; Smith & Squire, 2005).

In the present experiment efforts were made to at least partially conceal the shared size associations across trained stimuli to avoid this eventuality, for example through the use of filler pairs and the pseudo random exposure to premises in training. Despite these precautions the factor of heightened awareness cannot be discounted as contributing to the overall speeding of RTs observed here. Indeed, future investigations of this type would benefit from some form of explicit awareness measure regarding the relationships across premises (see Werchan & Gómez, 2013), particularly as the sleep state has been noted elsewhere to enhance the declarative awareness of hidden patterns and rules (Diekelmann, Wilhelm, Wagner, & Born, 2013; Stickgold & Walker, 2004; Wagner et al., 2004).

3.4.2 Sleep mechanisms and premise memory

Despite the above noted issue of integration, there still remains a provisional degree of evidence for consolidation in the present experiment. Analysis of premise pair performance indicates that the serial position effect was expressed equivalently at all three time delays, as evidenced by the consistent patterns of superior performance for end, relative to central, premises. While this is a null result in terms of the original experimental hypothesis, it should also be noted that RTs and accuracy levels did not diminish after a period of sleep relative to the other two time delays. This indicates that memory of premises learned immediately were equivalently retained across periods of wakefulness and sleep. While an appropriate degree of caution is required when interpreting this negative result, it can be argued that the present measures potentially reflect the maintenance declarative memory across time (Walker, 2005).

Supporting evidence for this claim derives from the weak association between measures relating to the serial position effect and sleep physiology. While no relationships between these measures arose at the sleep stage level, a more fine grained analyses relating to spindle activity found a predictive relationship between SWS spindle density and the extent that accuracy levels ran counter to the serial position effect (i.e. became higher for central, relative to end, items). Although this correlation did not survive correction for multiple comparisons, the predicted association between these specific measures and SWS spindles should also not be overlooked (Cox et al., 2012; Wilhelm et al., 2011). Furthermore, as the integrative measures in this experiment appear to exceed a threshold sensitive to consolidation any resulting relationships, even if genuine, may appear in a weakened form. This latter contention is perhaps evidenced by the lower levels of spindle activity in the present investigation relative to other word learning studies (cf. Clemens et al., 2005; Schabus et al., 2004; Tamminen et al., 2010).

In Chapter 1 it was emphasised that consolidation processes are not restricted to the sleep state. While there is ample evidence that sleep may provide a unique benefit to certain forms of consolidation (Drosopoulos, Schulze, et al., 2007; Ellenbogen, Payne, et al., 2006; Wagner et al., 2004) many investigations are unable to truly seperate this account from a passive-protective interpretation, where the sleep state provides minimal interference for consolidation processes operating over wakefulness (Muller & Pilzecker, 1900; Stickgold, 2009; Wixted, 2004). Indeed, in the literature wakefulness has been observed to facilitate the integration of novel word forms into a pre-existing mental lexicon both in parallel (Tamminen et al., 2010) and independently of sleep when training is spaced across the waking day (Lindsay & Gaskell, 2012). In this light, the present observation that *both* sleep and wakefulness may contribute to the maintenance of information does not necessarily diverge from the existing literature, and indeed provides some overlap with the results reported in Experiment 1, as accuracy levels for the majority of the integrative measures were similarly maintained across the two experimental time points.

3.4.3 Sleep and the maintenance of transitive inference performance

The arguments regarding an involvement of sleep in the maintenance of memory can also be extended to the TI measures. The ability to form TIs between elements of novel information (e.g. the inference ‘swock > slerk’), the critical measure of relational integration in the present experiment, did not differ across the three delay types in terms of accuracy or accurate RT performance. However, memory performance for these inferences similarly did not decay across these time periods. As such, a tentative case may also be put forward for a consolidation-associated maintenance of learned materials, particularly given the association between novel TI performance and sleep architecture. However, as with premise memory, this interpretation is limited by the negative findings on this measure, and is only raised here because of the expressed relationships between TI accuracy and sleep physiology.

In accordance with a maintenance account there was a positive relationship between SWS spindle activity and novel word TI accuracy. This suggests that, despite the superficially equivalent accuracy levels across three delay types, active consolidation processes during sleep may still support the flexible expression of recently learned novel information. Indeed, given the involvement of spindles specific to the SWS stage, these findings potentially indicate the involvement of systems-level consolidation as defined by the active consolidation model (Frankland & Bontempi, 2005; Gais et al., 2002; Rasch & Born, 2013). This observation is therefore provisionally compatible with the previously noted association between relational memory and SWS (Lau et al., 2010), as well as findings more broadly relating SWS and its affiliated spindle activity in the consolidation of declarative memories deemed to be of future relevance (Diekelmann et al., 2013; Wilhelm et al., 2013).

Accuracy for TIs involving real words (e.g. ‘sheep > lizard’) also displayed a positive relationship with sleep spindle activity, albeit with spindle density in stage 2 sleep rather than SWS. Although real word inferences formed a baseline comparison with regards to TI performance, the association between this measure and aspects of sleep architecture is not entirely surprising. Indeed, one of the most consistently replicated findings with regards to declarative memory consolidation is the relationship between NREM sleep spindle activity (in stage 2, SWS or both) and the memory for paired associate word learning, particularly as measured through cued recall (Clemens et al., 2005; Gais et al., 2002; Saletin et al., 2011; Schmidt et al., 2006). Although real word TIs were comprised of items not paired together in training, exposure to both items within the same size context may have similarly a cued their association, analogous to these previous studies.

3.4.4 Size generalisations

The final measure featured in the present experiment related to the potential emergence of the semantic distance effect for size judgments between trained novel words items and untrained but familiar word items (e.g. ‘swock ? worm’). It was predicted that sleep would enhance the expression of this effect for novel word size judgments, potentially through the enhanced imagability of novel word items due to their assimilation into semantic memory (Moyer, 1973; Pavio, 1975). Moreover, this effect was not predicted to occur for a newly introduced baseline comparison featuring size judgments between two real words (e.g. ‘zebra ? worm’) as these items already exist in semantic memory. Lastly, it was predicted that an emergent semantic distance effect for novel word items would correlate with one or more aspects of sleep physiology.

Taken together, the present findings do not confirm these predictions. Instead a rapid emergence of the semantic distance effect appears to be present for both types of size judgement (i.e. baseline and novel judgements) 20 minutes after training which persisted over 12 hours of sleep or wakefulness. Baseline real word vs. real word size judgments produced a robust semantic distance effect at all three time points, as expressed by faster RTs and higher accuracy levels for more distant, relative to proximal, size judgments (Moyer, 1973; Pavio, 1975). The size judgments between novel word and real word items similarly displayed evidence of the semantic distance effect, although performance was generally slower and less accurate relative to the baseline measures as might be expected. Critically, there was no differential impact of time delay on the strength of the semantic distance effect for novel word size judgments. Although there was a weak positive association between the magnitude of the serial position effect and SWS, this relationship did not survive correction for multiple comparisons suggesting a minimal involvement of sleep-specific consolidation on these measures, both behaviourally, and physiologically.

3.4.5 Possible circadian influences

A final point needs to be raised about possible circadian influences across the time delay manipulation. Specifically, responses were slower in the sleep condition on two baseline measures: the real word comparisons for both size generalisation and TI responses. These findings not only run counter the original hypothesis, which does not predict a slowing of RTs after sleep, but also indicate the presence of a possible confounding factor in the sleep condition.

A long standing issue in the field of sleep and memory research is the influence of time of day effects on memory performance. This relates to natural fluctuations in participant arousal through a combination of homeostatic and circadian pressures, as part of a normal sleep-wake cycle (Borbély & Achermann, 1999). Heightened metabolic activity from the mid-afternoon to late evening, as part of the circadian acrophase, has been linked to a general tendency for faster but less accurate responding on a range of experimental tasks (Blake, 1967; Monk & Leng, 1982). Of particular relevance is a study conducted by Tilley and Warren (1983) who probed RT performance in a semantic classification task in the morning (9:00am), afternoon (2:00pm) or evening (8:00pm). RTs on this measure, which involved allocating word items as in-group or out-group members of specific semantic categories, were the slowest in the morning by ~100 msec relative to the other two conditions albeit with a non-significant trend towards more accurate responding. While the tasks in the Experiment 2 clearly differ from the categorisation measure used by Tilley and Warren (1983) the present findings similarly suggest that the expressed memory of pre-existing semantic relationships may sensitive to time of day effects.

In the present experiment precautions were taken to minimise potential circadian factors on memory performance. For example, testing in the sleep condition was delayed by ~35 minutes after arousal to avoid the non-specific cognitive detriments of sleep inertia (Borbély & Achermann, 1999). Moreover, a number of supplementary measures were obtained to monitor the potential influence of sleep quality and fatigue on performance. Despite the general homogeneity across the three time delays on these measures the slowing of RTs in the sleep condition remain best explained by circadian influences. Unfortunately, despite the overall speeding of responses in this experiment relative to Experiment 1 it remains difficult to assess the extent that this factor may have compromised the critical measures of sleep integration. Future studies might benefit from additional RT tasks unrelated to the principle measures of interest to provide a more sensitive gauge of circadian pressures on performance across the waking day (see Nishida & Walker, 2007). Indeed, Experiments 4 and 5 reported in this thesis implement such a task.

3.4.6 Conclusions

Experiment 2 failed to replicate the patterns of RT integration seen across 24 hours in Experiment 1. However, these null results do not necessarily count against the contention that sleep can support the integration of semantically compatible information. Instead the removal of the incongruent condition from the modified TI paradigm featured in Experiment 2 appears to have resulted in considerably faster RTs on the derived behavioural measures. As this acceleration was seen prior to a sleep opportunity the sensitivity of these measures to the influences of consolidation was most likely reduced.

Despite these null results there is provisional evidence that sleep can support the flexible expression of declarative memory over time. This was potentially reflected in the relationships expressed between SWS spindle activity and the measures that showed the greatest performance improvement over time in Experiment 1. Had the incongruent condition been featured in Experiment 2 we may well have seen more robust evidence for systems-level consolidation during sleep. However, as they stand, these associations with sleep physiology did not survive correction for multiple comparisons. Moreover, the presence of a potential circadian confound in the sleep condition, as evidenced in the slowing of baseline real word responses, adds an additional level of caution to this interpretation.

Chapter 4: The role of prior knowledge and consolidation in the recovery and integration of memory.

4.1 Introduction

Observations from Experiment 1 indicate that prior knowledge can help expedite the integration of novel information over time, an effect found to occur only when training was marked by a degree of semantic ambiguity, perhaps explaining the lack of integration in Experiment 2. The relative differences in integration observed across these two investigations might be contingent on memory reactivation during consolidation, a process that may bolster memory performance to levels that otherwise require the exposure to additional congruent information during wakefulness (Bell, Kawadri, Simone, & Wiseheart, 2013; Lindsay & Gaskell, 2012).

This notion is compatible with evidence that consolidation-associated processes may be particularly effective for memories that are retained at sub-optimal levels (Diekelmann et al., 2009; Göder & Born, 2012; Stickgold, 2009). Alongside integration, this aspect of consolidation may help reconstruct information extrapolated from everyday learning environments marked by the intermittent exposure to useful information. While the notion of what exactly constitutes a ‘weakly’ encoded memory trace is a complex one (Conte & Ficca, 2012) there is nevertheless good evidence that consolidation can help recover impoverished memories in a number of ways (Diekelmann et al., 2009). One such recuperative dimension relates to retroactive interference (RI), an effect noted to occur when memories are retroactively inhibited through subsequent exposure to interfering or competing materials (Wixted, 2004). Interestingly, there is a long standing observation that memory consolidation can help both protect and recover newly learned information from the detrimental influences of RI, and perhaps even interference in general (Abel & Bäuml, 2013; Drosopoulos, Schulze, et al., 2007; Ekstrand, 1967).

An adaptive benefit of memory recovery in this manner is that individuals are able to learn opportunistically throughout a waking day without penalty to the robustness of their initial learning (Fenn et al., 2003). However studies exploring this effect to date, as with the majority of the memory consolidation literature, feature declarative materials with little or no relevance to pre-existing memory. Given the findings of Experiments 1 and 2, the present investigation aims to explore how the semantic congruency (and thus the potential usefulness) of newly formed memories may influence their ongoing integration and recovery from RI across time.

4.1.1 Consolidation and the recuperation of memory

The notion of what comprises a weak memory trace is a broad one spanning a number of distinct factors including (but not limited to) encoding depth (Drosopoulos, Schulze, et al., 2007), participant skill level (Diekelmann et al., 2010), the semantic coherence of learned materials (van Kesteren et al., 2012), the relevancy of new information (Diekelmann et al., 2013; Wilhelm et al., 2011), and the detrimental effects of competing or interfering memories (Ellenbogen, Hulbert, et al., 2006). Despite the numerous factors associated with the concept of memory strength there is now a growing body of evidence demonstrating that a consolidation can enhance relatively impoverished memories under a range of circumstances.

For example, in the procedural domain it has been found that the sequence elements considered to be the most difficult to learn in a finger tapping task will manifest the largest performance improvements after a night of sleep. Moreover, the relative difficulty of this task was predictive of sleep spindle activity, considered a biomarker of brain function and plasticity, over the subsequent night (Kuriyama et al., 2004). In a parallel finding from the declarative literature, memory for abstract word pairs (e.g. ‘union-rate’), determined to be harder to learn than concrete word pairs (e.g. ‘brick-tree’), similarly displayed a selective enhancement after sleep (Schmidt et al., 2006). Furthermore, as with Kuriyama et al. (2004), the recall of these relatively demanding paired associates was linked with a greater expression of spindle activity over a preceding nap. In terms of direct encoding strength, it has been demonstrated that single lists of unrelated word pairs trained to a lower criterion (≥60%) will display selective enhancements in retention after sleep not seen for items trained to a higher level (≥90%) (Drosopoulos et al., 2007, Experiment 2). Lastly, the consolidation-associated enhancement of false memory rates, thought to be dependent on the abstraction of shared semantic features across themed word lists, have been found to be especially pronounced in participants with comparatively low skill levels on this task (Diekelmann et al., 2010).

At present the literature appears to be convergent on the notion that consolidation provides a useful ‘salvage’ function for declarative materials retained at sub-optimal levels. Indeed the majority of published studies featuring paired word associates (a standard declarative task) tend to train participants to lower thresholds (i.e. ~60%) with the aim of inducing a level of retention maximally sensitive to the effects of consolidation (Marshall et al., 2006; Payne et al., 2012; Plihal & Born, 1997; Wilson, Baran, Pace-Schott, Ivry, & Spencer, 2012). In a notable exception to this pattern of findings Tucker and Fishbein (2008) found that the amount of non-rapid eye movement (NREM) sleep obtained over a nap was only predictive of performance improvements on a range of declarative tasks for participants with relatively high skill on these measures. However, these seemingly anomalous findings were derived from training protocols that were interleaved with testing procedures, potentially leading to a strong selective priming of these materials in skilled participants (Stickgold & Walker, 2013). Moreover, asymptotic levels of retention may benefit from consolidation through synaptic downscaling processes associated with slow wave sleep (SWS) (Tononi & Cirelli, 2003, 2006), as opposed to the reconstructive processes inherent in systems-level consolidation (Diekelmann et al., 2009; Stickgold, 2009).

4.1.2 Consolidation and the recovery of memory from interference

One of the longest standing observations in the literature relating to declarative memory consolidation is that retained information can become more resistant to interference over time, particularly after a period of sleep (Ellenbogen et al., 2009; McGaugh, 2000; Muller & Pilzecker, 1900; Wixted, 2004). More recently, passive-protective accounts of consolidation have been supplanted by evidence implicating the active role of sleep in the recovery of memories that have been detrimentally influenced by RI. Recovery in this manner has been established at two levels: (i) in the resurgence of performance on tasks influenced by the nonspecific effects of waking interference, and; (ii) in the recovery of memories from more specific forms of interference created by directly competing elements of information.

With regards to the former phenomenon it has been shown that the ability of participants to generalise the perceptual features of synthetic speech tends to deteriorate over the course of a waking day, yet 'rebound' to pre-interference levels after a night of sleep (Fenn et al., 2003). Similarly, the progressive deterioration of response times on a visual texture discrimination task across wakefulness, thought to be induced by a saturation of information in primary visual cortex, can also be reversed after a nap (Mednick et al., 2002). Furthermore, in this latter case, the extent of performance recovery was contingent on the length of the intervening nap and, in particular, the levels of SWS obtained during this period. In both of these investigations it is plausible that performance recovery was facilitated by systems-level consolidation, which may have helped unburden saturated memory systems from their accrued neural loads (Fenn et al., 2003; Mednick et al., 2002).

Recovery from interference has also been noted for tasks that involve more overtly conflicting materials. This has been assessed most directly using the classic A-B, A-C list learning paradigm (Barnes & Underwood, 1959; Wixted, 2004). Here participants are trained on two lists of paired associates: a primary list (e.g. 'book – egg' or 'A-B') and, after a short break typically containing a filler task, a secondary list sharing the same stimulus but different response term (e.g. 'book – dog' or 'A-C'). It has been robustly demonstrated through several variations of this task that the acquisition of the second A-C list will detrimentally, and retroactively, interfere with recall of the primary A-B list associations (Barnes & Underwood, 1959; Underwood & Postman, 1960). Furthermore, this effect is not observed when the two lists contain different stimulus items (e.g. ‘A-B’ and ‘C-D’ lists)(Drosopoulos, Schulze, et al., 2007). Interestingly, the impairment of A-B associates appears to be a temporary phenomenon, and one that can be reversed through consolidation.

Several classic studies have found that the effects of RI diminish with the passage of time. It has been shown that the memory recall of paired associates (unrelated adjectives) trained across A-B and A-C lists will reveal strong RI effects after five hours, as seen in the superior memory performance for A-C list items (Underwood, 1948a, 1948b). However, when testing is delayed by 48 hours RI effects will diminish, resulting in equivalent recall levels between the A-B and A-C lists. However, despite the replicated observation that RI recovery is a time-dependent phenomenon (Abra, 1969; Ceraso & Henderson, 1965; Rothkopf, 1957) the precise mechanisms underlying this change to behaviour have remained somewhat speculatory (Postman & Underwood, 1973; Slamecka & Ceraso, 1960).

The suggestion that sleep may be a critical factor in this process were first demonstrated by Ekstrand (1967) who assessed participants ability to recall items from A-B and A-C word lists after an 8 hour period containing overnight sleep or wakefulness. It was found that although memory performance in the sleep condition improved generally, it was most pronounced for the primary A-B list associations, implicating differential effects of consolidation across the two word lists. However, Ekstrand and colleagues later rescinded this interpretation after considering the results of a follow up experiment (Ekstrand et al., 1971). Here, recall after list learning was tested after either 20 minutes of wakefulness or a 7 hour retention interval involving a period of sleep systematically disrupted during the rapid eye movement (REM) or SWS stages. Contrary to their original findings, a non-significant trend for recovery was observed after 20 minutes, leading Ekstrand et al. to conclude that the resurgence in A-B list memory was the product of explicit rehersal within in the first 20 minutes after training.

However, the experiment conducted by Ekstrand et al. (1971) had numerous methodological issues, not least of which was the lack of a condition featuring an uninterrupted night of sleep. Deeming their conclusions premature Drosopoulos et al. (2007) set about exploring RI recovery using a number of suitable control conditions including list recall conducted immediately after learning, after a 20 minute delay, and after an overnight period of uninterrupted sleep or wakefulness. The key finding of this investigation confirmed that A-B list recovery emerged only after a night of sleep. This observation was recently expanded upon in a study conducted by Abel and Bäuml (2013) who adapted A-B, A-C list training protocols so that exposure to these associations induced both RI and proactive interference (PI), where the learning of primary associations (A-B items) proactively interfered with the retention of the secondary list (A-C items). It was found that both RI and PI displayed recovery exclusively after a night of sleep, not only re-confirming the observation that this state can aid in the recovery of RI, but also that these recuperative effects may extend to interference in general.

While the literature presently implicates sleep-associated consolidation in the active recovery of memories from interference it is important to note that for this process to be considered adaptive it must necessarily be accompanied by a degree of selectivity (Rasch & Born, 2013). Firstly, as previously indicated, our everyday learning environments are likely to feature the intermittent exposure to information that varies considerably in terms of its applicability to future behaviour (Wixted, 2004). A second important point is that if all impoverished memories were to be recovered without due differentiation then our resource limited memory systems might quickly become overwhelmed with redundant, and potentially maladaptive, information (Stickgold & Walker, 2013; Tononi & Cirelli, 2003, 2006).

The consolidation-associated recovery of impoverished memories is therefore unlikely to display a linear relationship with decreasing memory strength. Instead, the retention of weakly encoded information, at a certain point, is likely to drop below a threshold sensitive to the recuperative processes noted above, thus allowing less essential information to be more readily forgotten (Stickgold, 2009). Despite the important notion of selectivity in relation to memory consolidation there is presently little research into how the relative usefulness of newly acquired information may impact RI recovery.

4.1.3 Experiment 3 aims and predictions

The two experiments in this chapter feature protocols based loosely on the A-B, A-C list learning paradigm to explore not only how the recovery of information may relate to its semantic congruency, but also how such a manipulation might influence the application of memories to new but familiar information. As with Experiment 1, the effects of consolidation after training will be probed either immediately (after 20 minutes) or after an extended delay (24 hours). Furthermore, provisional sleep measures will be obtained from participants using home sleep recording kits to explore the potential sleep-associated mechanisms underlying any changes to behaviour.

Congruency will be manipulated by varying the semantic relationships shared across the A-B and A-C word lists. Unlike previous iterations of the A-B, A-C paradigm, which typically feature word items devoid of any implied semantic relationships (see Figure 22 A for an example), here a size judgment component will be introduced to list learning, permitting the manipulation of size relationships established across the 'B' and 'C' lists. More specifically, the common 'A' stimulus featured in both lists will consist of novel words that vary in terms of their inferred size between adjacent real word items. In conjunction, the relative size of these word triplets will either be complementary (e.g. 'camel>gorby>duck' or 'B>A>C') or contradictory (e.g. 'camel < gorby<duck' or 'B<A<C') (see Figure 22, B and C respectively). The training of novel word sizes in this manner is designed to permit a useful (congruent) or less useful (incongruent) size relationship to emerge across the two lists in relation to prior knowledge.

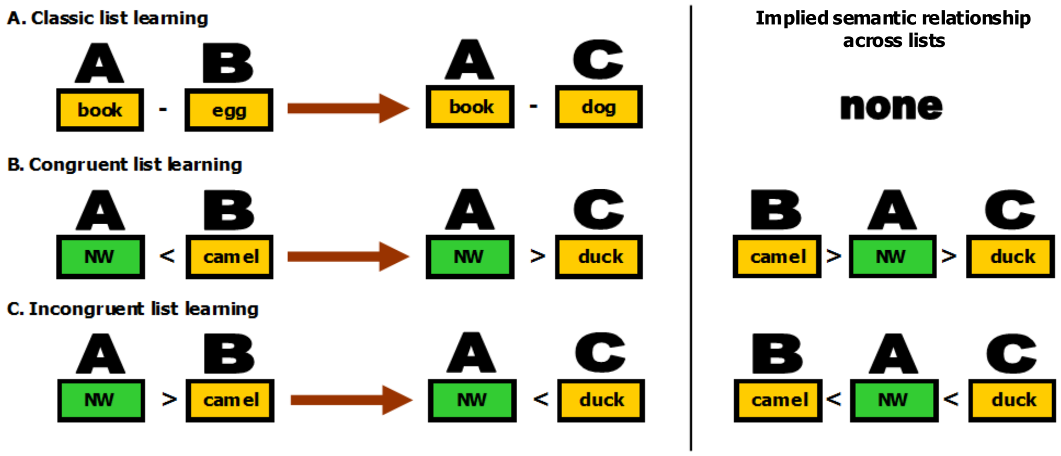


Figure 22. Examples of the A-B, A-C list learning paradigm. (A) A 'classic' version of the task in which no semantic relationship evident across the A-B and A-C lists. (B) An example of a congruent size relationship across trained lists. In this case the size of the novel word (NW) item can be inferred across lists (where '>' indicates 'bigger than' and '<' indicates 'smaller than') as determined by the size relationship shared with the two real word items in the 'B' and 'C' lists (camel and duck). (C) An example of an incongruent size relationship across trained lists. Here the NW is trained as being bigger than the larger item in the 'B' list (a camel) but smaller than the smaller item in the 'C' list (a duck) making the inferred size of the NW item ambiguous. The brown arrows in all three examples of this list learning paradigm represent the sequential order of list learning prior to testing.

Memory performance after a time delay will be assessed using three measures: cued recall to probe item memory, two-alternative forced choice (2AFC) size judgments to explore the memory of semantic relationships established across word lists and, lastly, a 2AFC generalization measure designed to assess how participants are structuring their generalisations of novel words to new, but familiar information.

List measures

The first two measures (cued recall and 2AFC list judgements) are where we might expect the differential effects of RI on list learning to be expressed. In the cued recall measure items will be cued by a common 'A' item (in this case a novel word) and participants will be requested to generate the affiliated 'B' and 'C' real word items. In the 2AFC list judgements, however, participants will be presented with two items (a real word and a novel word) and will be asked to select which item they perceive as being the largest.

As such these measures will probe item memory, and memory of the specific size relationships, respectively.

Here it is predicted that memory for A-B items will be selectively maintained after a consolidation opportunity while memory of the interfering A-C list will be reduced, as with the equivalent time points in previous investigations of this type (Drosopoulos, Schulze, et al., 2007; Underwood, 1948a, 1948b). Importantly, however, RI recovery may also manifest differentially as a function of list congruency. The relative compatibility of congruent triplets across the two lists may not only lead to their selective abstraction (Ellenbogen et al., 2007; Payne et al., 2009) but also an increased 'binding' of this information, akin to the unitisation effects observed for central premises in Experiment 1 (Walker & Stickgold, 2010). Should these word triplets be treated more as an independent unit by virtue of their linked semantic properties (e.g. 'B>A>C') we may expect to see A-B and A-C list performance levels rendered more comparable over 24 hours. Conversely, for incongruent information, we may expect RI effects to remain pronounced over time as these memories will benefit less from consolidation-associated integration.

It should also be noted that memory performance may differ depending on the type of measure used to assess list memory. Although the A-B, A-C paradigm typically probes memory through cued recall measures we may in fact expect RI effects to be more pronounced on the 2AFC list measure in this experiment, as it more directly relates to the relative size relationships established during training (as in Figure 22). Nevertheless, the cued recall measure may provide an interesting gauge of the errors made between lists. For example, we may expect to see an increased number of interlist errors for congruent items after a 24 hour delay, as the extrapolated relationships across B>A>C word triplets may render the context specific memory of 'B and 'C' items increasingly interchangeable across the two lists (Inostroza & Born, 2013; McClelland & Rogers, 2003).

Generalisation

The final behavioural measure will explore how participants are generalising novel word size relationships established across both lists during training. The size distances between the real word items presented in the 'B' and 'C' lists will systematically vary, ranging from smaller differences (e.g. 'rat>shusk>flea') to larger differences (e.g. 'horse>snafe>rabbit')(Figure 23, A). This differing 'focus' in the trained size distance between real word items may lead to inferred sizes for novel items that vary considerably in terms of their potential usefulness for subsequent generalisation. For example, novel items trained between real words with a more restrictive (narrow) size focus will potentially induce more accurate generalisations to a range of new, but familiar, real word items. Conversely, a broader (wider) size focus between real word items may be less useful in forming subsequent generalisations given the expanded range of potential size that these novel items may occupy (the green cones in Figure 23, B).

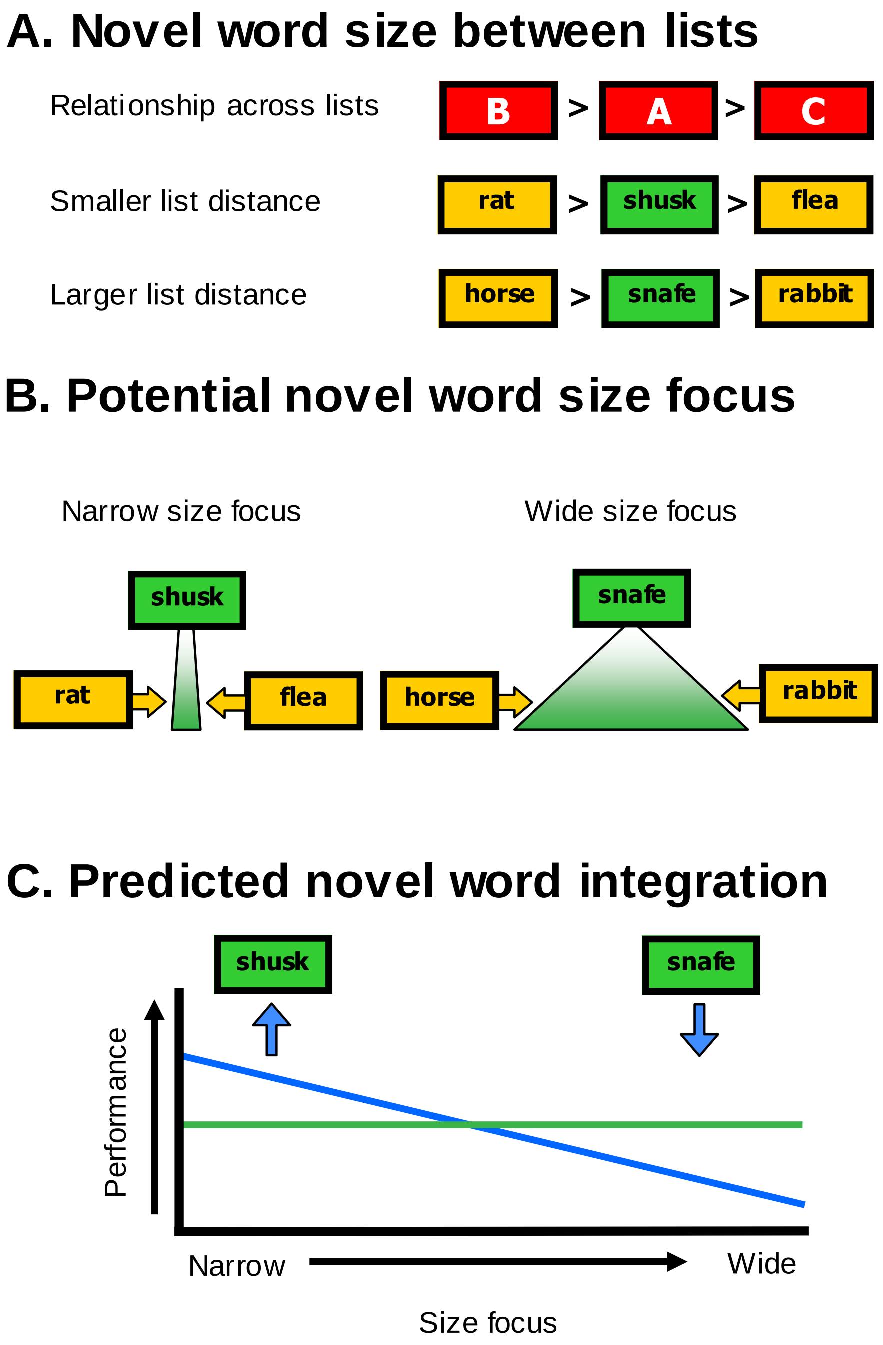


Figure 23. Examples of how the relative distances between real words may influence the subsequent generalisation of novel words to new but familiar real word items. (A) and (B) illustrate how smaller and larger size distances between real word list items may result in differing degrees of inferred size for the novel word items (the green cones). (C) Illustrates how these differences may influence subsequent generalisations. The green line in (C) represents performance prior to consolidation, with generalisation levels undifferentiated across the size focus dimension. The blue line and arrows indicate the potential direction of integration, with performance enhancements for novel words trained between list items with a narrower size focus, and a potential decrease in performance for novel words trained between real words with a larger size focus.

As a specific hypothesis it is predicted that novel words trained within a narrow size focus will display performance advantages relative to novel words trained within a wider size focus after a consolidation opportunity. In other words, the generalisation of novel information is predicted to more optimally reflect the underlying usefulness of the derived size relationships (the blue line in Figure 23, C). The reasons for this prediction are threefold. Firstly, previous research has shown that information with more applicability to future behaviour can be selectively consolidated (Stickgold & Walker, 2013; Wilhelm et al., 2011). Secondly, this effect may be particularly contingent on increased access to pre-existing semantic information through consolidation processes, as indicated by current theoretical (McClelland et al., 1995; McClelland & Rogers, 2003) and experimental accounts (Payne et al., 2009; Tse et al., 2007). Thirdly, should RI effects between the A-B and A-C list be reduced over time, as predicted above, then the potential size range of novel items may become more adequately 'bookended' by real word items, leading to more discerning generalisations on this measure.

Novel word generalisations will also be measured in the incongruent condition; however as accuracy measures cannot be directly determined for these items, responses will instead be coded in terms of their relationship to the upper size limits established in training. This measure (defined more in the following sections) will provide insight into the extent that participants are using trained size information to structure their subsequent generalisations of ambiguous novel words. The potential effects of consolidation on this measure are more speculatory. However, if consolidation facilitates performance in this condition then participants may increasingly form generalisations around the upper and lower size limits established during training after 24 hours. This may result in a similar pattern of generalisation performance seen in the congruent condition. Alternatively, if incongruous size relationships are more prone to decay (i.e. forgetting) we may expect lower levels of generalisation performance after 24 hours, with no differentiation between novel words with a narrow or wide trained size focus.

Sleep measures

Lastly, overnight sleep stage data will be obtained using home sleep recording kits. While these measures are not intended to replace the more stringent polysomnography (PSG) recordings featured in Experiment 2 they may nevertheless provide useful (albeit provisional) measures regarding any associations between sleep physiology and the above predicted effects.

Here, SWS is hypothesised to display one or more relationships with the above measures. Firstly, RI recovery may be contingent on the overnight reactivation of memories (Drosopoulos, Schulze, et al., 2007; Underwood & Postman, 1960). Given that mnemonic reactivation is proposed to occur predominately during in SWS we may find that these stages in particular are associated with an equalisation of performance between A-B and A-C lists (Diekelmann & Born, 2010a, 2010b). Furthermore, given the association between SWS sleep and semantic integration (Tamminen, Lambon Ralph, & Lewis, 2012) and generalisation (Payne et al., 2009) it is predicted that these stages may also relate to an enhanced capacity for the differentiation and generalisation of novel word sizes. Although we might also predict sleep spindle activity specific to the SWS stages as contributing to the above behavioural changes the home sleep recording kits used in this investigation are unfortunately not capable of registering these field potentials.

4.2 Experiment 3 Methods

Note that the methodology for Experiment 3 is very similar to that implemented in Experiments 4 and 5. As such, for brevity, these protocols will be outlined once in detail here with only deviations from these procedures reported and justified in the method sections of Experiments 4 and 5.

4.2.1 Participants

Thirty-three participants aged between 18 and 26 were recruited through the University of York Psychology electronic experiment booking system (PEEBS). All were healthy, non-smoking native English speakers with no prior history of sleep or psychiatric disorders. All participants agreed to abstain from caffeine, alcohol and napping over the course of the experiment and 24 hours beforehand. Written informed consent for participation was obtained in compliance with the local ethics committee.

Two participants were removed from analysis due to excessively slow RTs (>2.5 standard deviations) at both training and test. The remaining participants were randomly assigned to the immediate (n=15; 12 females, mean age = 20.3, *SD*=3.3) or delayed conditions (n=16; 15 females, mean age = 19.2, *SD*=.9). Participants had a mean age of 19.7 years (*SD*=1.8) and were provided payment or course credit for taking part, with an additional £20 Amazon gift voucher awarded to the best overall performer.

4.2.2 Design

The present experiment followed a mixed design featuring the between-participant variable of Delay type with testing administered 20 minutes (Immediate) or 24 hours (Delayed) after training (Figure 24, A). The within-participant variables include Congruency (Congruent or Incongruent), List (A-B list memory vs. A-C list memory) and the trained size focus of novel words (Narrow vs. Wide).

The training session consisted of three phases: (i) a short instructional phase followed by questions, (ii) a novel word familiarization phase, and; (iii) A-B and A-C list training through 2AFC trials (Figure 24, B). The testing session consisted of two phases: (i) cued recall of the word lists, and; (ii) a 2AFC size judgment task featuring both the original list items and series of new, but familiar real word comparisons. Training and testing sessions for all participants were staggered across the day (9am-6pm) to minimise time of day effects (Tilley & Warren, 1983).

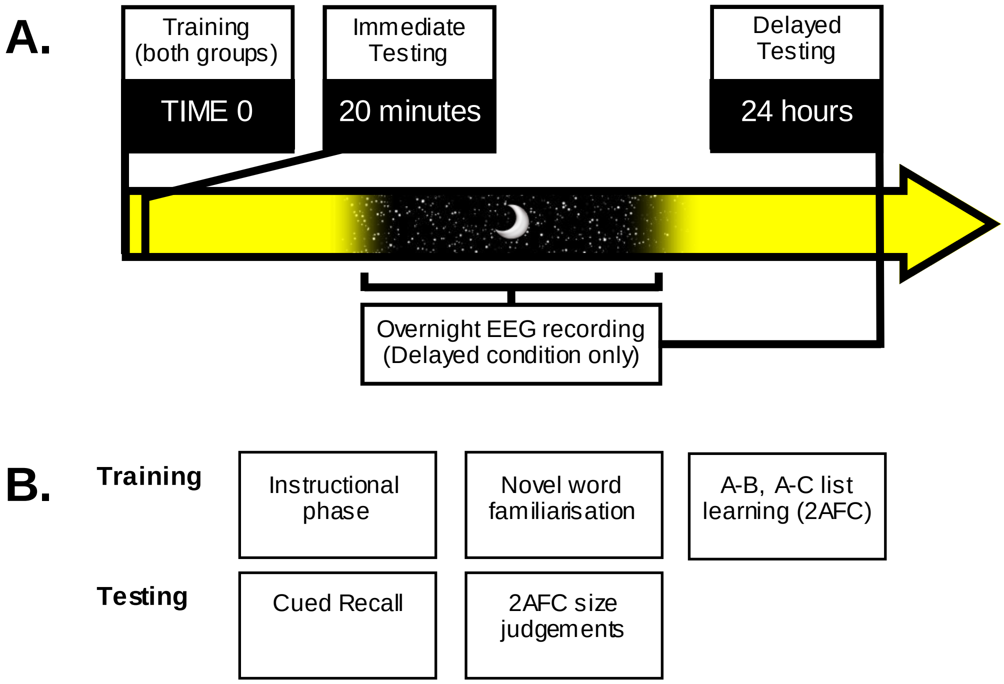


Figure 24. Experiment 3 design. (A) The between-participant conditions in this experiment. (B) The sequential order of tasks across training and testing.

4.2.3 Stimuli

Real words

As stimuli construction required significantly more rated animal items than those used in Experiments 1 and 2, animal items were drawn from Pavio's (1975) database of size-rated animal norms, with each animal item independently scored on a size 1-9 size scale (small-large) by 49 participants. Forty of these items were selected, ranging from 1.00 (flea) to 7.69 (giraffe), to feature across the 'B' and 'C' lists during training.

To counterbalance both congruency and list presentation two sets of ten real word pairs were created which were closely matched on five basic properties: (i) the average size of rated animals (*M*=4.1, *SD*=2.1), (ii) the variation in size distance between the two animal items (*M*=2.8, *SD*=1.5), (iii) bigram frequency (*M*=9304, *SD*=6239), (iv) word length (*M*=5.4, *SD*=1.9), syllable count (*M*=1.7, *SD*=.8), and; (v) word frequency in the English lexicon (*M*=14, *SD*=19.2). This latter factor was based on words per million as measured by the SUBTLEXus frequency database (http://subtlexus.lexique.org/). These stimuli sets and size ratings can be found in Table 16, while a more complete table of these items and their respective properties can be found in Appendix 7. Note that the total amount of trained A-

B, A-C word triplets in the present Experiment (n=20) is the same as previous investigations of this type (Drosopoulos, Schulze, et al., 2007; Ellenbogen et al., 2009).

Table 16. Stimuli sets and size ratings used for A-B, A-C list learning.

	Stimulus Set	B list	A novel item	C List	B list size rating	C list size Rating	Size difference (B-C)
	1	rat	shusk	flea	2.27	1.00	1.27
		wolf	brenk	cat	4.86	3.12	1.74
		quail	dewel	roach	3.02	1.20	1.82
		goat	plave	penguin	4.57	3.90	0.67
		cow	soble	eagle	6.47	3.82	2.65
		alligator	tenic	beaver	5.51	3.61	1.9
		donkey	badet	porcupine	5.92	3.53	2.39
		camel	gorby	duck	7.43	3.08	4.35
		rhinoceros	geter	dove	7.76	2.51	5.25
		elk	hesty	squirrel	6.33	2.61	3.72
	2	frog	smead	snail	1.98	1.20	0.78
		crab	hoxey	butterfly	1.98	1.35	0.63
		panther	hober	fox	5.63	3.98	1.65
		giraffe	bleem	tiger	7.69	5.90	1.79
		leopard	morax	mouse	5.59	1.98	3.61
		zebra	gatin	snake	6.22	3.02	3.2
		octopus	zoron	lobster	5.82	2.51	3.31
		moose	cherk	rooster	7.06	3.06	4
		bear	slend	bee	7.37	1.22	6.15
		horse	snafe	rabbit	6.65	3.06	3.59
Mean	1	N/A	N/A	N/A	5.41	2.84	2.58
	2	N/A	N/A	N/A	5.60	2.73	2.87
Grand Mean	Both	N/A	N/A	N/A	5.51	2.78	2.72

Note: Stimuli in this table represent the two training sets prior to counterbalancing. Size ratings were obtained from Pavio's (1975) database of size rated animal norms.

These two sets of real word items formed the basis of subsequent counterbalancing in terms of congruency and list appearance (i.e. presentation in the 'B' or 'C' lists). To ensure an equal number of 'bigger' and 'smaller' judgements across the congruency manipulation list counterbalancing was conducted for half the items in a given set (n=5) resulting in four possible combinations of trained stimuli, ensuring an equal balance of size judgements across congruency and list training combinations (Appendix 8).

Lastly, ten additional real word items were selected from Pavio's (1975) size rated norms to feature as untrained but familiar stimuli for the generalisation phase. These items were chosen to extend across the full spectrum of potential animal size, ranging from 'ant' (1.04) to 'elephant' (8.35) (Table 17). For the congruent portion of the generalisation phase accuracy measures for these items were formed on the basis of the upper and lower rated size limits established across training. For incongruent measures novel word size judgements were coded with regards to their relative size judgement in relation to the largest semantic item of a given word triplet.

Table 17. Experiment 3: real word animal items and ratings to be used in the generalisation phase.

Comparison item	Size rating
ant	1.04
moth	1.24
turtle	2.47
skunk	3.10
monkey	3.73
sheep	4.83
mule	5.98
lion	6.16
buffalo	7.27
elephant	8.35

Note: Stimuli ratings were obtained from Pavio's (1975) database of size rated animal norms.

Novel words

The 20 novel words featured as the common 'A' item across lists were created using Wordgen software (Duyck et al., 2004). These items were designed to be comparable to all 50 real word stimuli regarding their average word length (5 letters), bigram frequency (*M*=8826, *SD*=4128), and syllable count (*M*=1.7, *SD*=.05). All novel words were cross referenced with the Medical Research Council's (MRC) online psycholinguistic database to ensure their lack of existence in the contemporary English lexicon and then randomised into the subsequent construction of A-B, A-C word triplets using Excel (see Table 16).

4.2.4 Procedure

Excluding the primary instructional phase and cued recall measures, all experimental tasks were presented in a quiet testing room on a 17.1 inch computer monitor using the program Eprime 2 (Psychology Software tools, Inc., Pittsburgh, PA).

Instructions

As with Experiments 1 and 2, training was conducted under the pretence of the 'explorer scenario' to provide context for the training and testing tasks (see Appendix 3). However, several minor adjustments were made to these instructions in the present experiment.

Firstly, unlike Experiments 1 and 2, participants were informed that they would be tested on the information they had learned at a later point after training. Forewarning of testing in this manner may potentiate the relevancy of learned information for subsequent consolidation (Diekelmann et al., 2013; Wilhelm et al., 2011). Secondly, a justification for the presence of incongruent information was included prior to encountering the A-C list, with participants instructed that ‘new information has come to light’ and that relative animal sizes required ‘updating’ accordingly. In Experiment 1 incongruous information was covertly introduced in training, and was thus not expected by the participants, potentially leading to an increased variability in performance.

Training procedures

Instructional Phase

At time zero (Figure 24) participants read an instructional pamphlet clearly outlining the experimental protocols with various examples (Appendix 9) and contacted the experimenter once they felt they understood the demands of the experiment. Participants were encouraged to imagine the comparative size relationships between novel and real word items to render any potential mnemonic strategies across participants more comparable (Gais et al., 2002; Schabus et al., 2004; Schmidt et al., 2006). Once these instructions were understood the experimenter then asked the participants five standardised questions to ensure that they had attended to these instructions prior to training (~6 minutes) (Appendix 10).

Novel word familiarisation

To ensure prior attendance to the novel word forms participants then began a familiarisation phase where 20 novel words were presented three times in a randomised order as follows: (i) a blank 600 msec pause; (ii) a 600 msec fixation point; (iii) a 600 msec presentation of a novel word item; (iv) a 600 msec fixation point, followed by; (v) an untimed text field accompanied by a request for participants to type the word they saw, with instructions to press the ‘return’ key once completed. This cycle was repeated until 60 trials had been completed across a single block (~7 minutes). The experimenter remained present for the first ten trials to ensure that participants were comfortable with this task.

List learning

Participants were pseudorandomly allocated to one of the four counterbalanced training sets (as outlined in section 4.2.3). Each trial involved exposure to 20 real word-novel word pairs, the on screen location of which was counterbalanced (e.g. ‘A>B’ and ‘B<A’) to avoid a screen location bias, leading to a total of 40 trails per block. Participants made 2AFC size judgements between paired items using the left or right buttons of a ‘Cedrus’ model RB-520 button box corresponding to the side of the screen of the animal they thought was biggest. The order of presentation was as follows: (i) a blank 600 msec pause followed by; (ii) a 600 msec fixation point followed by; (iii) an untimed 2AFC judgement between paired items followed by; (iv) 1000 msec of feedback. This cycle

repeated until all 40 stimulus combinations had been randomly presented. As with familiarisation, the experimenter remained present for the first ten trials to ensure participants were comfortable with the task.

As with Experiments 1 and 2, feedback was provided for responses in the form of the words ‘correct’ or ‘incorrect’ presented in green or red text respectively, with accompanying accuracy scores (in %) at the end of each block. After the third training block a hidden exit criterion was activated (≥80%) which, if exceeded, would exit training once this block was completed. This exit criterion was introduced ensure that all participants had received approximately equivalent levels of training prior to testing. Seven opportunities were provided to exceed this threshold (10 blocks in total) before training terminated automatically regardless of performance on the final block.

Both A-B and A-C stimulus lists followed the above format (~12 minutes each). To minimise explicit rehearsal of the A-B list all participants took a mandatory 15 minute break between learning the two lists, where they played the computer game ‘Snood’ (www.snood.com) set at ‘medium’ difficulty. This simple puzzle game has been used previously as a filler task when exploring the effects of A-B, A-C inference (Drosopoulos, Schulze, et al., 2007).

Testing procedures

Once training was completed on the second A-C list participants were randomised into immediate (20 minute) or delayed (24 hour) conditions. Prior to this randomisation process, neither the experimenter nor participant was aware of the condition allocation. Participants in the 20 minute condition surfed the internet during this time, while participants in the delayed condition left the laboratory to continue their normal waking activities.

Cued recall

After the time delay participants were firstly asked to complete a paper-based cued recall measure in an A4 booklet containing all 20 novel word items in a fixed, randomised order (Appendix 11). Each novel word item had two blank spaces with a request to not only provide the two accompanying real word items, but also their list of origin (i.e. the first list or second list). This measure is analogous to the cued recall format used in the majority of previous RI investigations (Bower, Thompson-Schill, & Tulving, 1994; Drosopoulos, Schulze, Fischer, & Born, 2007; Ekstrand, 1967; Ekstrand, Sullivan, Parker, & West, 1971). This phase was untimed and took approximately ~11 minutes to complete.

Size judgements

The second phase of testing involved a series of 2AFC size judgements between: (i) the original 40 novel word-real word pairings presented across both A-B, and A-C lists at training, and; (ii) new generalisation comparisons between novel words and new, but familiar, real word items. Original and new size judgements were randomised and tested together to place demands on semantic flexibility and integration in this phase. The latter generalisation judgements required size comparisons between each of the 20 novel words in relation to the ten untrained animal items found in Table 17. In total 240 size judgements were presented in this phase twice in a counterbalanced fashion to avoid a screen location bias (480 trials in total) which were randomly presented across four testing blocks (120 items each).

As with training, participants were asked to choose the button representing the side of the screen that corresponded with the larger animal of a given pair. Each trial consisted of: (i) a blank 600 msec pause followed by; (ii) a 600 msec fixation point followed by; (iii) a timed 3000 msec forced choice between a novel word and real word followed by; (iv) a repetition of this cycle once 3000 msec had passed or the participant made a decision. All feedback and percentage performance cues were removed from this phase (~24 minutes in total).

Additional measures

As with all of the experiments presented in this thesis, the Stanford Sleepiness Scale (SSS) was used to assess participant alertness levels (Hoddes et al., 1973). This itemised measure required participants to rate their levels of alertness ranging from 1 (highly alert) to 7 (sleep onset) prior to training and testing. Participants were also asked to report their average sleep time (in hours) across the two weeks prior to the experiment as well as the amount of sleep obtained prior to the testing phase. For the immediate condition this latter measure referred to the night of sleep prior to coming into the lab, whereas for the delayed condition this referred to the night of sleep obtained between the training and testing phases. Lastly, participants were also asked to rate the quality of this sleep period on a 100 mm visual analogue scale (from worst to best).

Lastly, participants were screened for potential confounding issues using a series of prepared questions in a phone call including the requirements: (i) no history of mental health problems, (ii) no use of any psychiatric or mood altering medication, and (iii) English as a first language. These were later followed up by a paper questionnaire asking the same questions prior to the experiment.

4.2.5 Sleep recordings

Sleep patterns for participants in the delayed condition were recorded via wireless electroencephalography (EEG) derived from model 101 Zeo© home sleep recording kits (Zeo,Inc., Newton, MA, USA) (Figure 25). These products are commercially sold as alarm clock devices that awaken customers at 'optimal' points in their sleep cycle, however for the purposes of the present experiment this alarm feature was disabled. Participants who were allocated to the delayed condition were given a short instructional session (~5 minutes) on how to use this device after training was completed.

Figure 25. Experiment 3: The model 101 Zeo© home sleep recording device and wireless headband. Photo credit: © 2013 Zeo Inc. & Intus Healthcare Ltd.

EEG recordings were obtained using a lightweight wireless headband with three dry silver-coated fabric sensors with a topographical placement approximating the Fpz midline (Jasper, 1958). EEG signals were amplified from the headband to a bedside unit where sleep stages were automatically extracted online using a patented artificial neural network (Shambroom, Fabgregas, & Johnstone, 2012). When compared to the traditional Rechtschaffen and Kales (1968) sleep stage scoring criterion (as used in Experiment 2) output from the Zeo device similarly scores EEG recordings in 30 second epochs coded as 'light sleep' (stages 1 and 2 combined), 'deep sleep' (stages 3 and 4 combined), REM sleep, and wakefulness. Deep sleep, as identified by the Zeo, is therefore identical to the definition of SWS or stage N3 by the American Academy of Sleep Medicine (Schulz, 2008).

Three important caveats should be made about the use of Zeo devices in the present experiment. Firstly, the patented artificial neural network that automatically scores records is proprietary and therefore not available for examination by researchers. Secondly, although the Zeo device has produced favourable comparisons to manually scored PSG sleep recordings, many of these findings have been reported by researchers working for, or sponsored by, the manufacturers of the product (Scullin, 2012; Shambroom et al., 2012; Wright, Johnstone, Fabregas, & Shabroom, 2008). Comparisons with standard PSG recordings conducted by researchers without a conflict of interests have generally validated the Zeo with an overall agreement with expert human scorers of 81%, although the device was considered less consistent when scoring wakefulness, and displayed a tendency to underestimate the amounts of SWS (Griessenberger, Heib, Kunz, Hoedlmoser, & Schabus, 2012). Given the Zeo equipment's relatively low presence in the published literature thus far, any derived results should be treated with the appropriate degree of caution.

A third and equally important issue with the Zeo device relates to participant dropout, primarily due to the inadvertent removal of the electrode headband during sleep. While some studies report low levels of dropout for the device (Scullin, 2012; Shambroom et al., 2012) impartial research indicates large dropout rates approximating 43% (Griessenberger et al., 2012). In the present experiment five recordings were lost due to headband removal leading to a dropout rate of 29%, thus reducing the power of any subsequent analyses.

4.3 Experiment 3 results

All data were analysed using SPSS version 19.0 with a two-tailed $p<.05$ threshold for significance. All reaction time (RT) results are reported in milliseconds (msec). Extreme fast (<300 msec) RTs were removed from the data prior to analysis (1.2%).

4.3.1 Group equivalency

Training

To ensure the two lists were learned equivalently across the experimental conditions the numbers of blocks required to exit training (between 4-10) were entered into a 2x2 mixed-effects ANOVA using the between-participant variable of Delay type (Immediate vs. Delayed) and the within-participant variable of List (A-B vs. A-C).

No main effect of Delay type was present, $F<1$ (Immediate = 4.2, *SE*=.150; Delayed= 4.3, *SE*=.109) nor was there a main effect of List, $F<1$, (A-B = 4.3, *SE*=.15; A-C = 4.2, *SE*=.12). These factors were qualified by a marginal interaction between Group and List, $F(1,29)=2.97$, $p=.095$, driven by slightly fewer blocks required to exit training in the A-B list in the Immediate condition (*M*= 4.1, *SE*=.21) compared to the delayed condition (*M*=4.4, *SE*=.2). However, given the marginal nature of this interaction, and the lack of differences relating to performance at criterion exit, alertness levels and sleep measures (as below), this difference should be considered negligible.

Accuracy levels were compared on the final block of training (≥80%) via a mixed-effects ANOVA featuring the between-participant variable of Delay type (Immediate vs. Delayed) and the within-participant variables of Congruency (Congruent vs. Incongruent) and List (A-B vs. A-C). No main effects of Delay type, $F(1,29)=1.28$, $p=.268$, or Congruency, *F*<1, were apparent however a main effect of List emerged, *F*(1,29)=21.1, *p*<.001. Visual inspection of training performance in Table 18 indicates that the secondary A-C list accuracy was higher in all conditions, suggestive of practice effects after A-B list training. No interactions were present, *Fs*<1.06, *ps*>.311, confirming that accuracy levels were comparable between the Immediate and Delayed conditions prior to the testing phase.

Table 18. Experiment 3: Exit block performance.

Congruency	List type	Immediate condition	Delayed condition
Congruent	A-B	91% (2.3)	87% (2.3)
	A-C	97% (1.4)	94% (1.3)
Incongruent	A-B	90% (2.1)	89% (2.1)
	A-C	95% (1.8)	95% (1.7)

Note: Values in parentheses denote the standard error.

Self-reported sleep measures

To assess the relative amounts of self-reported sleep both obtained normally (i.e. the average amount obtained over the two weeks prior to the experiment) and on the experimental night (i.e. prior to the testing phase) a 2x2 mixed-effects ANOVA was conducted, featuring the between-participant variable of Delay type (Immediate vs. Delayed), and the within-participant variable of Sleep period (normal night vs. experimental night) using the dependent variable of sleep amount (in hours).

No main effect of Delay type was present, *F*(1,29)=1.5, *p*=.235 (Immediate= 7.7 hours, *SE*=.25; Delayed = 8.1 hours, *SE*=.24) nor was there a main effect of Sleep period, *F*<1 (Normal night=8 hours, *SE*= .18; Experimental night=7.9 hours, *SE*=.23). As such sleep patterns between the two Delay types appear to be equivalent, both in terms of their general consistency, and the night of sleep obtained prior to testing. Furthermore, the lack of interaction between Delay type and Sleep period, *F*<1, indicates that the sleep obtained on the experimental night did not differ significantly from the regular sleep patterns for either Delay type.

To measure sleep quality subjective ratings of the night of sleep obtained prior to testing on the visual analogue scale (0-10 cm, worst to best) was assessed using an independent samples t-test between Delay types. In accordance with the above analysis these scores did not differ between the Immediate (*M*= 6.4, *SE*=2.4) and Delayed conditions (*M*=5.7, *SE*=2.03), *t*<1.

Alertness

Lastly, alertness levels throughout the experiment were compared between the Immediate and Delayed conditions using the self-reported measures obtained on the SSS (Hoddes et al., 1973). A 2x2 mixed-effects ANOVA featuring the between-participant variable of Delay type (Immediate vs. Delayed) and the within-participant variable of SSS ratings prior to a given experimental Phase (Training vs. Testing) using SSS ratings (1-7, most alert – least alert) as the dependent variable.

These data, which can be found in Table 19, revealed a main effect of Phase, *F*(1,29)=5.42, *p*=.027, with participants rating themselves less alert prior to testing that training, most likely due to relative levels of fatigue accrued across training (the immediate condition) or a full day of wakefulness (the delayed condition). However, as there was no main effect of Delay type, *F*<1, or interaction between Delay type and Phase, *F*<1, alertness levels should be considered equivalent between conditions. In general participants rated themselves as moderately alert, but not wide awake, on this measure (Hoddes et al., 1973).

Table 19. Experiment 3: Stanford Sleepiness Scale (SSS) measures.

Phase	Immediate	Delayed
Training	2.2 (0.26)	2.3 (0.11)
Testing	2.6 (0.22)	2.6 (0.21)

Note: Values in parentheses indicate the standard error.

4.3.2 Cued recall performance across A-B and A-C lists

The first of three behavioural measures featured in this experiment relates to cued recall, where participants were cued with a novel word (e.g. ‘shusk’) and asked to allocate the two corresponding real word items in terms of list order (i.e. ‘B-list’ or ‘C-list’, which could be ‘rat’ and ‘flea’ respectively). The % recall used as the dependent variable in the analyses of this measure represent the % of total possible responses for a given response permutation.

All possible responses from participants on this measure were coded in terms of list allocation and novel word affiliation. Using examples drawn from hypothetical training on stimulus set 1 in Table 16, coded responses included: (i) no response, (ii) untrained intrusions (e.g. ‘badger’), (ii) incorrect list recall and incorrect novel word recall (e.g. incorrectly recalling ‘rat’ in the C-List with the incorrect novel word ‘brenk’), (iii) correct list recall but incorrect novel word recall (e.g. correctly recalling ‘rat’ in the B-list with the incorrect novel word ‘brenk’), (iv) incorrect list recall but correct novel word recall (e.g. incorrectly recalling ‘rat’ in the C-list with the correct novel word ‘shusk’), and (iv) correct list recall with correct novel word recall (e.g. the most optimally correct response, with ‘rat’ correctly recalled in the B-list with the correct novel word ‘shusk’). As with previous studies of this type coding in this manner allows for an examination of the types of recall independently or in combination (Abel & Bäuml, 2013; Drosopoulos, Schulze, et al., 2007).

Recall for correct and incorrect list responses

As the effects of context (and interlist conflation) are of particular interest to this investigation, recall was examined independently for real word items allocated to the correct list or incorrect list (regardless of novel word affiliation). Analysis of the former coded responses is therefore directly comparable to the reports of list recall % obtained from previous investigations of this type (cf. Abel & Bäuml, 2013; Drosopoulos, Schulze, et al., 2007) whereas analysis of the latter coded responses will provide information recall % recall in terms of interlist conflation. It should be noted analysis was also conducted independently for the individual response codings on this measure (as above) however, as these analyses revealed little of interest beyond the composite measures reported below, they are not discussed further.

Correct list recall

Real words coded as being recalled in the correct list (regardless of novel word affiliation) were combined and entered into a 2x2x2 mixed-effects ANOVA featuring the between-participant variable of Delay type (Immediate vs. Delayed) and the within-participant variables of List (B list vs. C list) and Congruency (Congruent vs. Incongruent).

These results, which can be found in

Table 20, revealed a main effect of Delay type *F*(1, 29)=5.02, *p*=.033, η_p^2=.148, indicating that the average recall of words across the eight possible recall list and congruency permutations was higher in the Immediate condition (*M*=52%, *SE*=4.25) relative to the Delayed condition (*M*=39%, *SE*=4.11). However, no main effect of List was apparent, *F*<1, nor was there a main effect of Congruency, *F*<1. Furthermore, no interactions were evident between any combinations of the three factors, *All Fs*<1. As such, correct list performance revealed higher levels of context-specific recall in the immediate condition as opposed to the delayed condition.

Table 20. Experiment 3: Recall rates for correct and incorrect list responses and untrained intrusion..

		Immediate		Delayed	
Analysis type	List	Congruent	Incongruent	Congruent	Incongruent
Correct list responses	A-B	55% (5.6)	51% (5.2)	39% (5.5)	36% (6.5)
	A-C	51% (6.8)	51% (6.8)	39% (5)	40% (6.5)
Incorrect list responses	A-B	12% (2.9)	13% (3.3)	17% (3.2)	13% (3.2)
	A-C	12% (3.2)	14% (4.3)	16% (3.2)	16% (3.2)
Untrained intrusions	A-B	6% (2.9)	4% (3.1)	13% (2.8)	8% (3.1)
	A-C	3% (3.2)	3% (3.4)	9% (3)	11% (3.3)

Note: Values in parentheses represent the standard error.

Incorrect list recall

Real words coded as being recalled in the correct list (regardless of novel word affiliation) were combined and entered into the same 2x2x2 mixed-effects ANOVA above. No main effects of Delay type, List or Congruency were apparent, *F*s<1. Furthermore, no interactions between Delay type and List, *F*(1,29)=2.28, Delay type and Congruency, *F*<1, Congruency and List, *F*<1, or three-way interaction between these variables, *F*<1, was evident. As such analysis of incorrect list responses, counter to predictions, revealed little to distinguish any of the three variables across a 24 hour delay with regards to interlist conflation (see Table 20).

Intrusions

The recall rate of untrained intrusions were of particular interest to this investigation as they may provide one measure of thematic abstraction (Gallo, 2010; Payne et al., 2009). Accordingly, these responses were entered into the same 2x2x2 ANOVA as above, and are reported in Table 20.

This analysis revealed a marginal main effect of Delay type with a higher percentage of untrained intrusions in the Delayed condition (*M*=10%, *SE*=2.59) than the Immediate condition (*M*=4%, *SE*=2.67), *F*(1,29)=3.19, *p*=.085, η_p^2=.099, indicating that participants in the Delayed condition were more prone to generating untrained animal items on this measure. No main effect of List, *F*(1,29)=1.43, *p*=.241, Congruency, *F*<1, or any interactions between the three variables in this analysis was present, *All Fs*<2.6, *All Ps*>.128.

Interestingly, when taken together, these results indicate that the factors of List and Congruency had little impact on participant recall for correct and incorrect lists. At first blush this outcome is perhaps unusual given that RI effects in the literature are typically expressed on this measure (Baddeley & Dale, 1966; Bower et al., 1994; Underwood, 1948a, 1948b). However, this observation does not forgo an influence of RI and consolidation on the specific size relationships established in training, as reported in the following section. Moreover, while these findings indicate a diminished capacity for context-specific recall after 24 hours, this observation was not paralleled by the predicted increase in interlist conflation across the same time period. Instead, participants in the delayed condition displayed an increased tendency to generate more untrained animal word items on this measure. It is interesting to note that when absolute levels of recall are compared between the two time delays (regardless of list, congruency or intrusions), no differences emerge between the Immediate (*M*=68% *SE*=5.34) and Delayed (*M*=64%, *SE*=5.17) conditions, *t*<1, suggesting that the above findings reflect a qualitative shift in the type of recall in this condition, rather than absolute diminished levels of recall after 24 hours.

4.3.3 2AFC size judgements and retroactive interference

The second measure regarding list memory relates to the relative sizes of these items established during training for congruent (e.g. 'camel > gorby > duck') and incongruent (e.g. 'camel < gorby < duck') word triplets. Accuracy levels for 2AFC size judgements on this task were entered into a 2x2x2 mixed-effects ANOVA featuring the between-participant variable of Delay type (Immediate vs. Delayed) and the within-participants variables of Congruency (Congruent vs. Incongruent) and List (A-B vs. A-C).

Accuracy levels

This analysis yielded a main effect of Delay type, *F*(1,29)=4.28, *p*=.048, indicating that accuracy was generally lower in the Delayed condition (*M*=59%, *SE*=1.8) relative to the Immediate condition (*M*=64%, *SE*=1.67). A large main effect of Congruency was also evident, *F*(1,29)=101.26, *p*<.001, revealing a substantial drop in accuracy levels in the Incongruent condition (*M*=52%, *SE*=1.65) relative to the Congruent condition (*M*=71%, *SE*=1.53). The magnitude of this effect indicates that the presence of incongruous size relationships substantially compromised the ability to accurately recall the relative size relationships in this condition at both experimental time points. Furthermore, a main effect of List was apparent, *F*(1, 29)=23.01, *p*<.001, with A-B list accuracy (*M*=57%, *SE*=1.57) being generally lower than that of A-C accuracy (*M*=66%, *SE*=1.6), suggestive of RI effects. Importantly, these findings were qualified by a Delay type x List interaction, *F*(1, 29)=7.9, *p*=.009, η_p^2=.214, which is illustrated in Figure 26.

This interaction appears to be driven by the RI reduction across both congruency conditions after 24 hours. A series of paired samples t-tests within the Immediate condition reveal superior A-C list performance relative to the A-B list for both the Congruent, *t*(14)=-4.36, *p*=.001, and Incongruent responses, *t*(14)=-2.62, *p*=.02, strongly indicative of RI effects. However, this difference between list performance disappears for responses made in both the Congruent, *t*(15)=-1.1, *p*=.287, and Incongruent conditions, *t*<1, after a 24 hour delay. As such the overall interaction on this measure was driven by a reduction in the influence of RI after a consolidation opportunity. No additional interactions were observed between Delay type and Congruency, *F*<1, Congruency and List, *F*<1, or all three of these factors, *F*<1.

Overall these results indicate that, despite the considerable impact of Congruency on size judgement performance for both Delay types, the effects of RI were similarly minimized in both of these conditions over a 24 hour period. As predicted, accuracy levels for the interfering A-C list were reduced to levels approximating those of the A-B list after a delay, albeit for both Congruent and Incongruent information. Given the differential impact of the time delay on list accuracy, these results strongly indicate an influence of consolidation in the preservation of A-B list size relationships (Drosopoulos, Windau, et al., 2007; Ellenbogen, Payne, et al., 2006).

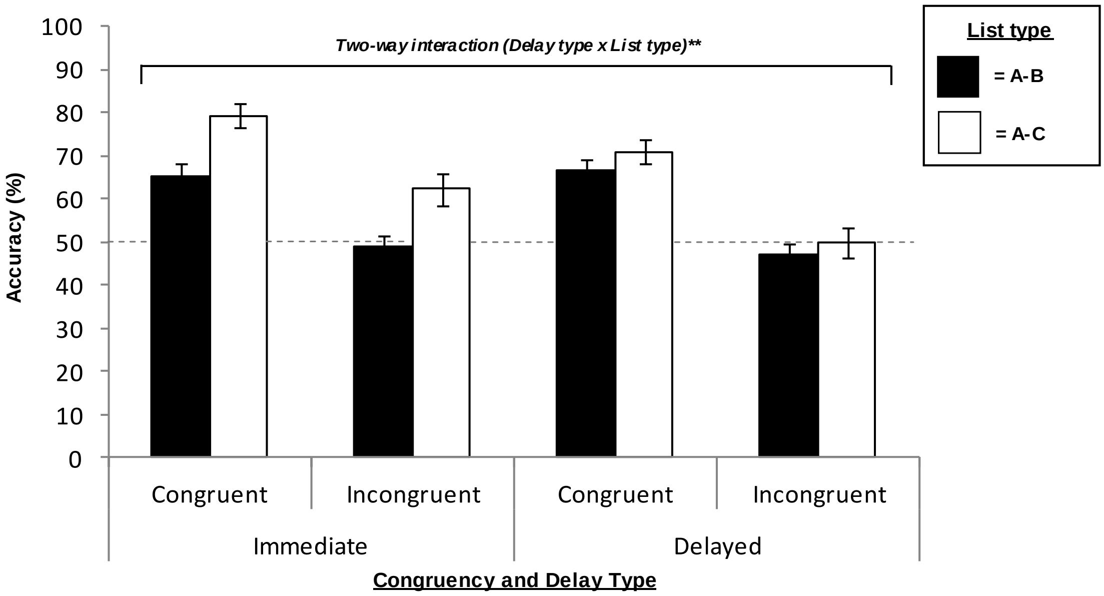


Figure 26. Experiment 3: 2AFC accuracy levels at test across Congruency, List type and Delay type. Significance value (**) indicates *p*<.01. The grey dashed line in this figure indicates chance levels of responding (50%). Error bars represent the standard error.

An important caveat regarding the above interaction is that performance was only at above chance levels in the Incongruent condition for Immediate A-C list items, *t*(14)=3.46, *p*=.004, while Immediate A-B items, *t*<1, Delayed A-B items, *t*(15)=-1.56, *p*=.140, and Delayed A-C items, *t*<1, were all at chance levels on this measure (the grey dashed line in Figure 26). As such, while congruency appears to be sufficient but not necessary for RI reduction, only congruent information appears to have been retained across the time delay. This latter observation suggests that the above noted reduction in RI effects are also part of a more global reduction in the ability to remember incongruent size relationships across time.

RT performance

Given the chance levels of incongruent accuracy noted above RTs, as opposed to accurate RTs, were analysed for 2AFC list judgements. As such, RT measures were entered into the same 2x2x2 (Delay type x Congruency x List) mixed-effects ANOVA as above, and the products of this analysis can be found in Table 21.

Table 21. Experiment 3: 2AFC reaction time measures across Congruency and List type and Delay type.

	Immediate		Delayed	
	Congruent	Incongruent	Congruent	Incongruent
A-B List	1279 (65)	1310 (61)	1342 (64)	1331 (60)
A-C List	1226 (55)	1248 (63)	1273 (53)	1292 (61)

Note: Values in parentheses indicate the standard error.

There was a large main effect of List, *F*(1,29)=11.21, *p*=.002, induced by the relative slowing of RTs for A-B list judgements (*M*=1316 msec, *SE*=41) relative to the A-C list (*M*=1260 msec, *SE*=39), in line with the overall patterns of RI as observed above. No main effects were observed for Delay type, *F*<1, or Congruency, *F*<1, nor were any interactions found between Delay type and Congruency, *F*<1, Delay type and List, *F*<1, List type and Congruency, *F*<1 or a three-way interaction between these factors, *F*<1. As such, despite the interesting effects observed for the accuracy measures, there was little to distinguish RT performance across these factors.

4.3.4 2AFC generalisations and novel word size focus

The third and final behavioural measure in this experiment relates to the generalisation of novel word items to new, but unfamiliar real word items. For the congruent condition the approximate accuracy of performance on this measure could be determined on the basis of a given novel word's hypothetical size limits established during training. However, in the incongruent condition no such 'correct' response could be determined as novel words were effectively trained as occupying two conflicting sizes. Nevertheless, responses in this latter condition can still provide useful information regarding the way generalisations are being performed. Accordingly, incongruous size judgements that were within (i.e. smaller) than the upper size limit of trained responses were allocated a positive % value, and items judged as exceeding this upper size threshold (i.e. larger) were allocated a negative % response value. As the coding of congruent and incongruent responses is analogous these measures will be analysed together however the dependent variable 'accuracy %' will be relabelled as 'generalisation %' to acknowledge the fact that there are no correct responses in the incongruent condition.

Of particular interest to the present experiment is how the trained size focus of novel words might influence the way generalisations are formed over time. For example, in the congruent condition the inferred size of novel words will vary along a continuum, with some items trained between two real words that share a narrow range of potential size (e.g. 'rat>shusk>flea') while others will occupy a wider range of potential size (e.g. 'horse>snafe>rabbit'). Accordingly, for analysis a factor of 'Size focus' was created by performing a median split along the size dimension between real word items present in the B-A-C word triplets (ranging from .6 to 6.15). Novel words trained between real words with a more concentrated size range (ranging from .6 to 2.39) were designated as 'Narrow', while novel words trained between real words with a broader size range (ranging from 2.65 to 6.15) were designated as 'Wide'.

Generalisation % analysis

The factor of Size focus was entered into a 2x2x2 mixed-effects ANOVA featuring the between-participants variable of Delay type (Immediate vs. Delayed) and the within-participant variables of Congruency (Congruent vs. Incongruent) and Size focus (Narrow vs. Wide). The dependent variable in this analysis, which can be found in Table 22, is generalisation % as defined above.

Table 22. Experiment 3: Generalisation performance for across 24 hours.

		Immediate		Delayed	
Congruency	Size focus	Generalisation (%)	RTs (msec)	Generalisation (%)	RTs (msec)
Congruent	Narrow	66 (2.5)	1173 (61)	69 (2.4)	1194 (59)
	Wide	67 (2.1)	1170 (57)	60 (2)	1191 (54)
Incongruent	Narrow	70 (2.4)	1165 (58)	67 (2.3)	1207 (56)
	Wide	71 (2.1)	1169 (58)	65 (2.1)	1165 (56)

Note: Narrow and Wide focus items are based on a median split of the rated size distance between two real words (see text). Values in parentheses represent the standard error.

No main effect of Delay type emerged, *F*(1,29)=2.66, *p*=.113, nor was there a main effect of Size focus, *F*(1,29)=1.4, *p*=.246, however a large main effect of Congruency was apparent, *F*(1,29)=7.82, *p*=.009. This main effect was driven by lower levels of generalisation % in the Congruent (*M*=68%, *SE*=1.02) relative to the Incongruent condition (*M*=65%, *SE*=1.19). A marginal Delay type x Size focus interaction was also present on this measure, *F*(1,29)=2.93, *p*=.098, η_p^2=.092. This interaction derives from a tendency for participants in the Immediate condition to produce similar generalisation levels for both Narrow (*M*=68%, *SE*= 2.17) and Wide focus responses (*M*=69% *SE*=1.7) whereas in the Delayed condition Narrow responses (*M*=68%, *SE*=2.1) were notably higher than Wide responses (*M*=62%, *SE*=1.69). Although only a trend, this marginal interaction is in the direction we might expect if participants in the Delayed condition were increasingly forming generalisations around the trained Size focus of novel items. No interactions were found between Delay type and Congruency, *F*(1,29)=1.15, *p*=.292, Congruency by Size focus, *F*(1,29)=1.62, or three-way interaction between the above factors, *F*(1,29)=1.34, *p*=.256.

To supplement the above ANOVA the factor of Size focus was broken down into its component units of trained size distance (ranging from .06 to 6.15) which were then independently correlated with generalisation (%) performance for the Immediate and Delayed conditions using Pearson's correlation coefficients. Interestingly, for Congruent responses (seen in the top half of Figure 27) a strong negative relationship between these variables was evident in the Delayed condition only, *r*(160)=-.325, *p*<.001, while generalisation performance remained relatively homogenous across the Size focus units in the Immediate condition, *r*(150)=-.006, *p*=.945. A Fisher's r-to-z transformation confirmed that the magnitude of this relationship was significantly different between the two time delays, Z=-2.89, *p*<.001. This observation is in line with the prediction that generalisations will increasingly relate to novel word size focus after 24 hours. In the Incongruent condition (the bottom half of Figure 27) no such relationship was expressed in either the Immediate, *r*(150)=-.002, *p*=.983, or Delayed conditions, *r*(160)=-.024, *p*=.760. Moreover, unlike the Congruent condition, a Fisher's r-to-z comparison did not reveal a significant difference in the magnitude of these correlations over 24 hours, Z=.19, *p*=.849.

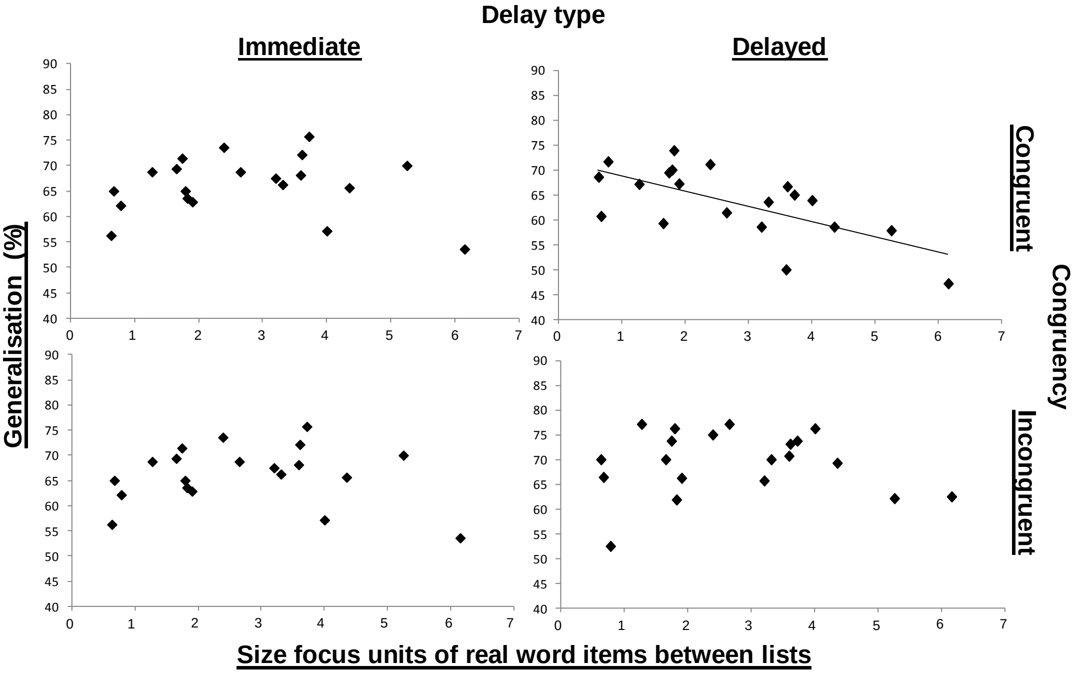


Figure 27. Experiment 3: Generalisation levels (%) correlated with the differences in size focus between real word items for both congruent and incongruent responses across the 24 hour time delay. Note that generalisation scores in this figure are averaged across each size focus unit.

Generalisation RT analysis

RTs, as opposed to accurate RTs, were entered into the same 2x2x2 mixed-effects ANOVA as above. The results of this analysis can also be found in Table 22. No main effect of Delay type was present, *F*<1, nor was there a main effect of Congruency, *F*<1, or Size focus, *F*(1,29)=1.37, *p*=.251. No interactions were found between Delay type and Size focus, *F*(1,29)=1.5, *p*=.231, Delay type and Congruency, *F*<1, Congruency and Size focus, *F*<1, or a three-way interaction between the above factors, *F*(1,29)=1.26. As such, there was little to distinguish any of the variables of interest on this measure.

Taken together it would appear that, in line with the initial predictions, the trained Size focus of novel words will increasingly influence the way generalisations are formed after 24 hours. This was demonstrated in the higher levels of generalisation performance for novel words with an increasingly restrictive size focus after an extended delay. Moreover, the impact of this factor appears to be present for congruently trained size relationships only. The lack of differences across time for incongruent generalisations runs counter to the original experimental predictions, as there was no evidence that these size judgments were more closely structured around prior knowledge across time. Nevertheless, the above chance levels of performance in this condition at both time points demonstrates that participants are still cable of forming approximate incongruent generalisations within the size limits established during training. Overall, the behavioural outcomes of this measure indicate that the usefulness (i.e. size focus) of B-A-C word triplets will only be reflected in congruent generalisation performance after an extended delay, suggesting the potential integration of these size relationships into semantic memory.

4.3.5 Sleep measures

Table 23 shows the main sleep parameters derived from the 11 intact recordings in Experiment 3. Participants slept for an average of 7.6 hours, with a mean sleep onset latency of 25 minutes. These sleep parameters are comparable to those expected from a sample of young healthy adults (Spriggs, 2008, p. 15). Additional comparisons with a large corpus of sleep data (Ohayon, Carskadon, Guilleminault, & Vitiello, 2004) corroborates the approximate levels of age-appropriate % sleep expected for TST (~450 mins), REM sleep (~22%) and 'Deep sleep' (20%) although levels of 'Light sleep' (~60%) may be slightly underestimated.

Table 23. Experiment 3: Sleep parameters for participants in the Delayed condition (n=11).

Sleep parameter	Mean time (min) ± SEM	% of total sleep time ± SEM
Total sleep time	456 ± 20.7	
Sleep latency	25 ± 3.3	5 ± 0.7
Nocturnal wakefulness	6 ± 1.9	1 ± 0.5
REM sleep	129 ± 12.6	28 ± 1.8
Light sleep (Stages 1 and 2)	241 ± 9.6	53 ± 1.6
Deep sleep (Stages 3 and 4)	87 ± 6.8	19 ± 1.3

Note: Sleep parameters in parentheses represent the relative sleep stages as defined by Rechtschaffen and Kales (1968), REM = rapid eye movement, SEM = standard error of the mean.

Behavioural factors and multiple comparisons

Following the rules established for correlational analysis in this thesis outlined in the previous chapter the behavioural measures obtained in this experiment were simplified as much as possible to represent the principle effects of interest.

For the cued recall and 2AFC list measures the amount of RI between the B and C lists is of particular relevance to this investigation. Accordingly, A-C recall was deducted from A-B recall to produce a composite value representing the degree of interlist competition. This factor was calculated independently for Congruent and Incongruent responses across both correct and incorrect forms of list recall. This factor was similarly calculated for the accuracy and RTs obtained from the 2AFC list measures. A total of eight factors were therefore generated for correlational analysis.

For 2AFC generalisation measures the principle effect of interest was the differential impact of size focus on generalisation performance. As such, the factor of Size focus (Narrow or Wide) was reduced down to a single variable representing the divergence in performance on these measures. This was done by deducting Narrow generalisation performance from Wide generalisation performance to produce a single composite value representing the extent of this difference. This factor was calculated independently across the congruency manipulation for both % performance and RTs, producing an additional four factors for correlational analysis.

With regards to sleep physiology, four measures of interest were extracted from the Zeo scored sleep data: total sleep time (TST, minutes), REM sleep %, 'light' sleep % (stages 1 and 2), and SWS % (stages 3 and 4). Sleep stages were calculated in terms of % of total sleep time, to control for the absolute amount of time spent in these stages. Given the non-standardised approach to multiple corrections in the sleep literature bonferroni corrections, as with Experiment 2, were established by multiplying the alpha value (set at .05) by the total number of comparison factors (n=12).

Correlational analysis

The corresponding table for this analysis can be found in Appendix 12. Despite the evidence suggesting a potential influence of consolidation on all three of the behavioural measures featured in this experiment only two provisional relationships emerged in relation to sleep physiology, neither of which survived correction for multiple comparisons.

Firstly, greater RI reductions in the measures of Congruent recall were associated with a larger percentage of time spent in REM sleep, $r(11)=.640$, $p=.034$ (uncorrected). This effect, although not predicted, is in the direction we might expect if aspects of REM sleep physiology were in some way reducing the interference effects for semantically congruent information (Norman et al., 2005; Walker & Stickgold, 2010). Secondly, the reduction of RI effects for Congruent 2AFC list comparisons was positively associated with TST, $r(11)=.665$, $p=.025$ (uncorrected), this latter correlation might reflect a non-specific benefit of sleep on list recovery for congruent items, and is at the very least in the direction we might expect if sleep was involved in memory consolidation. However, the significance value of the latter relationship should be considered particularly tenuous, as the removal of one participant with particularly prominent A-B list recovery in relation to the A-C list (+29%) rendered this relationship non-significant, $r(10)=.443$, $p=.199$, perhaps evincing the lack of power in this analysis.

4.4 Experiment 3 summary and discussion

The aim of Experiment 3 was to explore how the semantic compatibility of information across word triplets in the A-B, A-C list learning paradigm might influence RI recovery and semantic integration over time. This experiment yielded three main findings indicative of potential consolidation-associated behavioural changes in: (i) the de-contextualisation of context-specific memory as measured by cued recall, (ii) the selective maintenance for A-B list size relationships as measured by 2AFC list judgements, and (iii) an impact of novel word size training on generalisation performance after 24 hours when B-A-C word triplets shared congruent, but not incongruent, size relationships. All three of these key findings are not directly accountable by differences in baseline sleep patterns, alertness or training performance.

Cued recall

With regards to cued recall the first finding of note is the general lack of congruency and RI effects on this measure. The negligible impact of these factors is particularly striking when contrasted with their strong influence on the memory of the size relationships established across word lists. However, this outcome should not be considered anomalous given that the cued recall measures do not directly relate to the size relationships formed during training. Furthermore, although testing was expected by participants, the format of cued recall was not, perhaps further shifting participant's attention (and retention) towards the relative size relationships established across the B-A-C word triplets. The lower degree of correct list recall in this experiment (between 55 and 36%) compared to other studies featuring a similar training criterion (typically ~80% as in Drosopoulos, Schulze, et al., 2007) is perhaps testament to eventuality.

However, despite the above caveats, differences between the two time delays were observed on the cued recall measure. Namely, participants recalled fewer items in the correct context (i.e. list) and generated marginally more untrained intrusions relating to the semantic theme (e.g. animals) after a 24 hour delay. This qualitative shift in performance did not influence the absolute amount of words recalled in either condition, which remained approximately equivalent at both time points (~66%). The lack of congruency and RI effects on this measure, in conjunction with a decrease in list specific recall potentially implicate the passive forgetting of context-specific memory over time (Wixted, 2004), rather than the consolidation-associated abstraction of semantic information, as originally predicted.

That said, elsewhere in the literature reductions in context-specific recall have been attributed to the adaptive influence of consolidation. For example, experiments featuring the Deese–Roediger–McDermott (DRM) paradigm have found that although the veridical memory of semantically themed word lists can deteriorate over time (e.g. 'bed', 'night', 'dream') untrained word items maximally associated with the semantic theme of a given list (e.g. 'sleep') can be simultaneously enhanced after a period of sleep (Payne et al., 2009). Here, an active role for memory consolidation was argued on the basis of a negative association between veridical recall and SWS, suggesting the active suppression of context-specific memory details during this sleep stage. The de-contextualisation of semantic information in this manner may be a useful outcome of consolidation as the 'gist' of information can extrapolated while simultaneously reducing the need to retain potentially superfluous details (Brainerd & Reyna, 2002). Taken together, the shift away from context-specific recall in conjunction with the increased rate of untrained intrusions in the present experiment may be an extension of this effect, leading to an enhanced capacity for generalisation after a delay.

A-B, A-C list size judgements

Memory for the size relationships established across the word lists produced an extremely interesting set of results. Despite the profound impact of incongruent training, which resulted in drastically diminished memory for trained size relationships at both time points, RI effects had dissipated for both congruent and incongruent information after a 24 hour delay. The large drop in A-C list accuracy driving this effect is particularly noteworthy given that A-C list training performance was actually superior to the A-B list for all participants. Furthermore, the differential impact of a time delay on list performance strongly suggests an active role for consolidation, as a passive-protective account of sleep would implicate a similar reduction in performance across time in the memory for both the A-B and A-C lists (Drosopoulos, Schulze, et al., 2007; Ellenbogen, Payne, et al., 2006).

However, these effects require additional clarification. In the predictions made for this measure it was hypothesised that RI recovery would be most pronounced for congruent size relationships, possibly due to the increased binding of word triplets across lists by virtue of their complementary semantic relationships (e.g. 'B>A>C'). However, the above noted effects implicate a reduction of RI effects for *both* incongruent and congruent list materials. As such, it might be argued that although semantically congruent information is sufficient to induce the reduction of RI effects over time, it does not appear to be necessary.

That said, the memory for incongruent list relationships partly driving this interaction were compromised to the extent that performance for both A-B and A-C list items were statistically indistinguishable from chance (i.e. ~50%) after 24 hours. This latter observation may indicate a second tier of mnemonic selectivity nested within primary RI reductions observed on this measure, with memory for the less useful semantic relationships being suppressed or forgotten entirely over time (Wixted, 2004). Whether this effect is derived from the passive decay (i.e. forgetting) or active suppression of the Incongruous A-C list in particular, remains to be seen, and is explored more thoroughly in the following chapter.

Novel word size generalisations

The generalisation of congruent novel word items increasingly reflected the usefulness of congruent B-A-C size relationships after a consolidation opportunity. Specifically, a 24 hour time delay resulted in higher generalisation levels for novel words trained with a narrow (and more specific) size focus while generalisation performance for novel words trained with a wider (and less specific) size focus diminished. Conversely, responses remained relatively homogenous across the size focus dimension 20 minutes after training. Interestingly, these effects occurred without influencing the absolute levels of generalisation accuracy across the two time delays highlighting, as with the cued recall measures, a qualitative shift in the expression of memory after 24 hours.

In this instance the evidence more strongly indicates the potential integration of novel word size relationships into semantic memory over time, as performance after a delay was contingent on the usefulness of the size dimension established during training. As such, this observation is provisionally compatible with the notion that systems-level consolidation across 24 hours can facilitate the interaction between recently acquired memories and pre-existing knowledge (McClelland et al., 1995; McClelland & Rogers, 2003). Moreover, this finding suggests that an extended time delay may confer the selective enhancement of new information that has more relevance to future behaviour (Stickgold & Walker, 2013).

However, the time delay manipulation did not differentially influence the way incongruent generalisations were formed, nor did the relative size focus of novel words in this condition. Interestingly, however, participants were still able to form approximate novel word generalisations within the size constraints established during training. This was evidenced by size judgements that consistently fell within the upper (e.g. gorby<giraffe) and lower (e.g. gorby>duck) size range across B-A-C word triplets. Indeed, this overall pattern of generalisation was actually higher than that observed in the congruent condition. However, the main effect of congruency in this case appears to be driven by the impact of size focus in congruent performance, as reflected in the less accurate responses for wide focus items.

Sleep data

Despite the evidence for potential consolidation-associated changes to memory on all three of the behavioural measures featured in this experiment the predicted association between memory performance and SWS did not emerge. Instead two relatively weak relationships were found involving REM sleep and TST respectively, which did not survive correction for multiple comparisons.

Firstly, a positive association was observed between congruent RI reduction on the cued recall measure and an increase in the percentage of REM sleep, despite the lack of RI effects on this measure in the principle analysis. Although not predicted, this finding is provisionally compatible with modelling work conducted by Norman et al. (2005), who propose that theta oscillations in REM sleep may help dissipate RI effects by simultaneously strengthening weaker A-B list memories and suppressing stronger A-C list memories. However to fully corroborate this model this finding clearly requires replication.

The second relationship to emerge was a similar association between RI reduction across congruent A-B and A-C size relationships and TST. This latter finding suggests that sleep in general may aid in the dissipation of RI effects over time, in line with sequential models of memory consolidation which propose that the beneficial impact of the sleep on memory is contingent on the cyclic alternation between sleep stages as a whole (Ambrosini & Giuditta, 2001; Giuditta et al., 1995). Despite the lack of involvement of SWS the above relationships are therefore in a direction that we might expect if sleep was positively influencing congruent memory performance.

However, as noted when reporting these results, analysis of these relationships suffered from reduced power through the necessary exclusion of incomplete sleep records (see Appendix 21), an unfortunate drawback regarding the Zeo sleep recording device (Griessenberger et al., 2012). As they stand, the above relationships should be considered particularly tenuous given both the diminished number of contributing sleep records and the negligible significance of these relationships.

4.5 Experiment 4

As discussed in section 4.1.1 consolidation processes appear to be particularly beneficial in recovering partially fragmented, incomplete or impoverished memories. Although the four main findings uncovered in Experiment 3 were obtained after relatively few training blocks (n=~4) these effects nevertheless emerged after moderately high levels of training performance were reached (≥80%). However, it is not presently clear how training participants to near asymptotic levels (i.e. ≥95%) might influence the potential impact of consolidation on the three measures featured in Experiment 3.

A primary motivation for Experiment 4 is to see if RI recovery can truly extend to incongruent information. While the results from Experiment 3 indicate the reduction of RI effects for both congruent and incongruently trained word triplets across time, the chance levels of performance in the latter condition make it difficult to distinguish the effects of RI recovery from the absolute levels of interference induced by incongruency. If above chance A-B performance is observed for incongruent items at a higher training threshold it will provide evidence that RI recovery can extend to both forms of congruency (Drosopoulos, Schulze, et al., 2007).

A second motivation for Experiment 4 relates more broadly to the influence of encoding depth on consolidation. As highlighted in section 4.1.1 this factor has been found to influence the impact of consolidation on the retention of declarative materials. The effects of encoding depth were examined directly in one of the experiments conducted by Drosopoulos, Schulze, et al. (2007) who manipulated the training criterion for a set of non-conflicting word lists (e.g. A-B, and C-D lists) to a lower (≥60%) or higher (≥80%) threshold. It was found that overnight performance enhancements in recall were only observed when training was established at the former criterion, suggesting that more elaborative encoding can render declarative memories less sensitive to the influences of consolidation.

It was noted in Experiment 2 (Chapter 3) that mnemonic enhancement in this manner may be contingent of memory reactivation during sleep. For example, although the integration of novel word forms into a pre-existing lexicon typically requires a period of sleep (Dumay & Gaskell, 2007; Gaskell & Dumay, 2003; Tamminen, Davis, Merkx, & Rastle, 2012) it has been revealed that multiple training sessions spaced across the waking day can induce an equivalent level of lexical integration (Lindsay & Gaskell, 2012). As such, one benefit of declarative memory reactivation may be that it simulates the additional 'training' required to bolster performance to higher levels. If performance is already sufficiently high prior to consolidation, akin to the overexposure of congruent information in Experiment 2, then the recuperative mechanisms associated with consolidation may cease to become activated (Stickgold, 2009).

However, it is not presently clear whether higher levels of training will enhance or negate the behavioural effects found in Experiment 3. The potential influence of consolidation in Experiment 3 was noted despite relatively robust levels of training exceeding 80%, possibly because the detrimental impact of both RI and incongruency on list memory provided sufficient levels of mnemonic interference to engage memory consolidation. As such, it may be that the higher levels of training in Experiment 4 may serve to reinforce the congruent or incongruent nature of the list materials, resulting in a greater degree of selectivity when it comes to the subsequent integration of this information (Stickgold & Walker, 2013). In Experiment 4 these outcomes are explored by extending the training phase, thus increasing the depth of encoding of information prior to a consolidation opportunity.

4.6 Experiment 4 methods

The design, procedure and experimental instructions featured in Experiment 4 were virtually identical to those in Experiment 3 (section 4.2). As such, only the two deviations made to these protocols are outlined and justified here.

4.6.1 Participants

Thirty-six participants aged between 18 and 23 were recruited through the University of York Psychology electronic experiment booking system (PEEBS). All were healthy, non-smoking native English speakers with no prior history of sleep or psychiatric disorders. All participants agreed to abstain from caffeine, alcohol and napping over the course of the experiment and 24 hours beforehand. Written informed consent for participation was obtained in compliance with the local ethics committee.

Four participants were removed from analysis for failing to score ≥95% in training (n=3) and for a self-reported sleep disturbance on the experimental night (n=1). The remaining thirty-two participants were randomly assigned to the immediate (n=16; 9 females, mean age = 20.5, *SD*=3.2) and delayed conditions (n=16; 9 females, mean age = 19.3, *SD*=3.2). Overall, participants had a mean age of 19.9 years (*SD*=2.4) and were provided payment or course credit for taking part, with an additional £20 Amazon gift voucher awarded to the best overall performer.

4.6.2 Procedure

Reaction time task

A short 2AFC RT task was introduced prior to the training and testing phases to provide an objective measure of alertness to supplement the more subjective SSS measure (Hoddes et al., 1973). This task was adapted from an equivalent RT measure featured in an experiment by Nishida and Walker (2007), and follows a similar format to the stimuli presented in the primary 2AFC list learning and testing phases. Instead of word pairs participants were given a choice between the numbers '1' and '0' randomly presented 18 times in a counterbalanced fashion and were instructed to press the left or right buttons on the button box corresponding to the side of the screen containing the '1'. This task typically took less than a minute for participants to complete.

Average SSS scores shared a positive relationship with average RT task scores, *r*(32)=.421, *p*=.16. As such, the slower the RTs the less alert participants were rating themselves, suggesting that both measures are valid.

Training

More elaborative encoding in this experiment relative to Experiment 3 was assured in two ways. Firstly, the exit criterion after the third training block was removed, with all participants now required to complete a fixed set of six training blocks independently for both A-B and A-C list items. As the majority of participants completed training on the first block to include the exit criterion of ≥80% criterion in Experiment 3 (*M*=4.2, *SD*=.07) the expanded number of training blocks in Experiment 4 should allow performance to reach near ceiling levels. Secondly, participants were only included in analysis if their performance on this 6th block was ≥95% on both lists (hence the three removals noted above). These prolonged training protocols increased the time required to complete list training to ~32 minutes.

4.7 Experiment 4 results

All data were analysed using SPSS version 19.0 with a two-tailed *p*<.05 threshold for significance. All reaction time (RT) results are reported in milliseconds (msec). Extreme fast (<300 msec) RTs were removed from the data prior to analysis (0.7%).

4.7.1 Group equivalency

Training

Accuracy levels were compared on the final block of training using a mixed-effects ANOVA featuring the between-participants variable of Delay type (Immediate vs. Delayed) and the within-participants variables of Congruency (Congruent vs. Incongruent) and List (A-B vs. A-C). A main effect of Congruency emerged on this measure, *F*(1,30)=7.57, *p*=.010, with levels of Congruent accuracy (*M*=98%, *SE*=.49) exceeding those in the Incongruent condition (*M*=96%, *SE*=1.03). However, as there was no main effect of Delay type, *F*(1,30)=2.34, *p*=.124, List, *F*<1, or interaction between Delay type and Congruency, *F*<1, Delay type and List, *F*<1 or three-way interaction between these factors, *F*(1,30)=2.1, *p*=.158, training performance should be considered equivalent across the two time delays (Table 24).

Table 24. Experiment 4: Exit block performance at the end of training for both Congruency and List types.

Congruency	List type	Immediate condition	Delayed condition
Congruent	A-B	100% (.81)	99% (.81)
	A-C	99% (.82)	98% (.82)
Incongruent	A-B	97% (.81)	95% (1.8)
	A-C	98% (.82)	95% (1.5)

Note: Values in parentheses denote the standard error.

Self-reported sleep measures

To assess the relative levels of sleep obtained normally (i.e. over the two weeks prior to the experiment) and on the experimental night (i.e. prior to testing phase) a 2x2 mixed-effect ANOVA was conducted using the between-participants variable of Delay type and the within-participants variable of Sleep period (normal night vs. experimental night) using the dependent variable of self-rated sleep amount (in hours). No main effect of Delay type was apparent, *F*(1,30)=1.12, *p*=.298, nor was there a main effect of Sleep period, *F*<1. However, an interaction did emerge between these two factors, *F*(1,30)=6.02, *p*=.02. Visual inspection of Sleep ratings in Table 25 indicates that is interaction was driven predominantly by differences between the two Delay types in terms of regular sleep patterns, with participants in the Immediate condition reporting approximately 45 more minutes of regular sleep. Importantly, however, the amount of sleep obtained on the experiment night (i.e. prior to testing) was identical between the two conditions.

Table 25. Experiment 4: Self-rated sleep measures.

Sleep period	Immediate	Delayed
Normal night	8.2 (.21)	7.5 (.21)
Experimental night	7.8 (.24)	7.8 (.24)

Note: Sleep amount is reported in hours. Values in parentheses represent the standard error.

To measure subjective levels sleep quality ratings of the night of sleep obtained prior to testing the visual analogue scale (0-10 cm, worst to best) was assessed using an independent samples t-test between Delay types. A marginal difference did emerge between the Immediate (*M*= 5.3, *SE*=2.5) and Delayed conditions (*M*=6.8, *SE*=.24), *t*(26.57)=2.02, *p*=.054, indicating a tendency for the Delayed participants to rate their sleep on the experimental night of higher quality.

Subjective alertness levels

Alertness levels throughout the experiment were compared between the Immediate and Delayed conditions using the self-reported SSS measure. A 2x2 mixed-effects ANOVA was conducted using the between-participants variable of Delay type (Immediate vs. Delayed) and the within-participants variable of experimental Phase (Training vs. Testing). These alertness levels, which can be found in Table 26, revealed no main effect of Delay type, *F*(1,30)=1.15, *p*=.224, main effect of experimental Phase, *F*(1,30)=2.5, *p*=.124, or interaction between these two factors, *F*(1,30)=2.5, *p*=.124. As such the grand mean of all SSS ratings (*M*=2.3, *SE*=.12) indicated levels of alertness approximating 'Functioning at high levels, but not at peak; able to concentrate' for both delay types across training and testing (Hoddes et al., 1973).

Table 26. Experiment 4: Stanford Sleepiness Scale (SSS) measures at training and testing for the Immediate and Delayed conditions.

Phase	Immediate	Delayed
Training	1.9 (.21)	2.5 (.21)
Testing	2.4 (.21)	2.5 (.21)

Note: Values in parentheses indicate the standard error.

Objective alertness levels

Lastly, performance was compared on the newly introduced RT task. Accuracy and accurate RT measures were entered into a 2x2 mixed-effects ANOVA using the between-participant variable of Delay type (Immediate vs. Delayed) and the within-participant variable of experimental Phase (Training vs. Testing). The products of these analyses can be found in Table 27.

Table 27. Experiment 4: Accuracy (%) and Accurate RTs (msec) on the rapid response control task.

	Immediate		Delayed	
Phase	Accuracy (%)	Accurate RTs (msec)	Accuracy (%)	Accurate RTs (msec)
Training	99% (.56)	331 (9.9)	99% (.56)	341 (9.9)
Testing	98% (.85)	318 (7.1)	99% (.85)	311 (7.1)

Note: Values in parentheses indicate the standard error.

For accuracy levels no main effect of Delay type, *F*<1, Phase, *F*(1,30)=1.7, *p*=.202, or interaction between these two factors was observed, *F*<1. Analysis of the RTs for these accuracy levels also indicated no main effect of Delay type, *F*<1, although a main effect of experimental phase was evident, *F*(1,30)=12.43, *p*=.001, driven by faster RTs at testing than training. This difference most likely represents practice effects on this task. Importantly, no interaction between Delay type and Phase was evident, *F*(1,30)=2.06, *p*=.161, indicating equivalent RT performance across the two time delays.

4.7.2 Cued recall performance across A-B and A-C lists

As with Experiment 3 all cued recall responses were individually coded (see section 4.3.2). As analysis of the individual coded responses revealed little of interest beyond the two principle ANOVAs outlined below they are not reported here. To focus on the effects of context-specific recall on list performance responses were aggregated into measures of correct and incorrect list recall.

Correct list recall

Responses indicating correct list recall % were collapsed and entered into a 2x2x2 mixed-effects ANOVA featuring the between-participants variable of Delay type (Immediate vs. Delayed) and the within-participants variables of List (B list vs. C list) and Congruency (Congruent vs. Incongruent). The products of this analysis can be found in Table 28.

A strong main effect of Delay type was present, *F*(1,30)=17, *p*<.001, η_p^2=.362, with the average recall of correct lists items in the Immediate condition (*M*=67%, *SE*=3.5) exceeding that of the Delayed condition (*M*=46%, *SE*=3.5), thus replicating the equivalent finding in Experiment 3. No main effect of Congruency was apparent, *F*(1,30)=2.8, *p*=.105, nor was there a main effect of List, *F*<1. Furthermore, no interactions were found between Delay type and List, *F*<1, Delay type and Congruency, *F*<1, Congruency and List, *F*(1,30)=1.23, *p*=.277, or three-way interaction between these factors, *F*(1,30)=1.6, *p*=.215.

Table 28. Experiment 4: Recall rates for correct and incorrect list responses and untrained intrusion rates.

		Immediate		Delayed	
Analysis type	List	Congruent	Incongruent	Congruent	Incongruent
Correct list responses	A-B	64% (5.5)	68% (5.2)	44% (5.5)	53% (5.2)
	A-C	66% (4.4)	69% (4.3)	44% (4.4)	44% (4.3)
Incorrect list responses	A-B	8% (2.8)	5% (2.6)	12% (2.8)	23% (4.3)
	A-C	9% (4.3)	14% (3.5)	11% (2.6)	21% (3.5)
Untrained intrusions	A-B	3% (1.6)	3% (1.7)	4% (1.6)	5% (1.4)
	A-C	3% (1.4)	4% (1.7)	3% (1.7)	2% (1.7)

Note: Values in parentheses represent the standard error.

Incorrect list recall

Responses recalled in the incorrect list (%) were combined and entered into the same 2x2x2 ANOVA as above. A main effect of Delay type emerged, *F*(1,30)=6.08, *p*=.02, η_p^2=.168, driven by higher levels of incorrect list recall in the Delayed (*M*=17%, *SE*=2.2) condition relative to the Immediate condition (*M*=9%, *SE*=2.2). A main effect of List was also apparent, *F*(1,30)=13.06, *p*=.001, indicating increased levels of interlist conflation for the A-C list (*M*=17%, *SE*=2.2) relative to the A-B list (*M*=9%, *SE*=1.6). No main effects of Congruency were apparent, *F*<1, nor was there an interaction between Delay and Congruency, *F*<1, Delay and List, *F*(1,30)=1.45, *p*=.238, Congruency and List, *F*<1 or interaction between these three factors, *F*(1,30)=1.79, *p*=.192. These findings, which can also be found in Table 28, therefore reflect a higher degree of interlist conflation after 24 hours relative to 20 minutes.

Intrusions

The number of untrained intrusions (e.g. 'badger') was assessed independently. Recall % for these responses were entered into the same 2x2x2 (Delay type x List x Congruency) mixed-effects ANOVA as above. This analysis, seen in Table 28, revealed no main effect of Delay type, *F*<1, List, *F*<1, or Congruency, *F*<1. Furthermore there was no interaction between Delay type and List, *F*(1,30)=1.74, *p*=.197, Delay type and Congruency, *F*<1, Congruency and List, *F*<1, or interaction between the three factors, *F*<1. Consequently, untrained intrusion rates remained equivalent across the two time delays.

Taken together the cued recall measures indicate a drop in context-specific recall after a 24 hour delay, similar to the equivalent effect noted in the previous experiment. However higher levels of training appear to have simultaneously reduced the number of untrained intrusions, but increased the number of interlist errors, after an extended time delay. As with Experiment 3, the absolute % levels of recall (i.e. all coded responses) were compared between the two experimental time points. However in this case a difference did emerge across the time delay, with Immediate recall (*M*=79%, *SE*=3.87) exceeding that found in the Delayed condition (*M*=66%, *SE*=3.87), *t*(30)=2.31, *p*=.028. As such the above noted changes in memory recall appear to occur in parallel with an overall quantitative reduction in absolute levels of recall after 24 hours.

4.7.3 2AFC size judgements and retroactive interference

For this second measure of A-B, A-C list memory performance, accuracy levels at test were entered into a 2x2x2 mixed effects ANOVA featuring the between-participants variable of Delay type (Immediate vs. Delayed) and the within-participants variables of Congruency (Congruent vs. Incongruent) and List (A-B vs. A-C).

Accuracy levels

A main effect of Delay type emerged, *F*(1,30)=13.32, *p*=.001, with higher levels of accuracy in the Immediate condition (*M*=70%, *SE*=1.6) relative to the Delayed condition (*M*=62%, *SE*=1.6). A large main effect of Congruency was also present, *F*(1,30)=26.05, *p*<.001, reflecting higher levels of accuracy in the Congruent condition (*M*=73%, *SE*=1.7) compared to the Incongruent condition (*M*=59%, *SE*=1.82), however no main effect of List was apparent, *F*<1. These main effects were qualified by a significant Delay x List interaction, *F*(1,30)=9.6, *p*=.004, η_p^2=.242 (discussed below). No additional interactions were found for Delay type and Congruency, *F*<1, Congruency and List, *F*<1, or three-way interaction between these variables, *F*<1.

As with Experiment 3, the Delay x List interaction was driven by differential list performance across the two time delays, as illustrated in Figure 28. Namely, A-B list accuracy appears to be maintained across a 24 hour period, while A-C list accuracy levels are simultaneously reduced in both congruency conditions. Paired samples t-tests implemented between these lists revealed a non-significant trend towards RI effects in the Immediate condition, *t*(15)=-7.3, *p*=.104, whereas RI effects are present for Incongruent responses, *t*(15)=-2.14, *p*=.049. Despite the visible trend for superior A-B list performance relative to the A-C list after 24 hours these differences remained non-significant in both the Congruent *t*<1, and Incongruent conditions, *t*(15)=1.49, *p*=.158. Additionally, the only condition to display performance at chance levels were A-C responses in the Incongruent condition *t*<1, while incongruent A-B accuracy levels, unlike those seen in Experiment 3, were maintained at above chance levels, *t*(15)=2.3, *p*=.036.

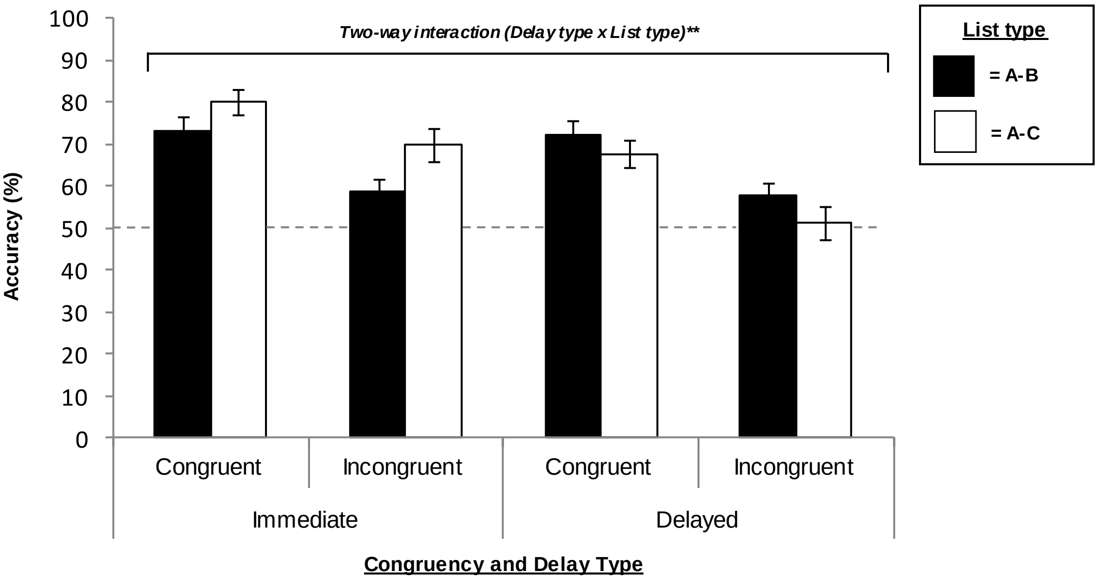


Figure 28. Experiment 4: 2AFC accuracy levels at test across Congruency, List type and Delay type. Significance value (**) indicates *p*<.01. The grey dashed line in this figure indicates chance levels (50%) of performance. Error bars represent the standard error.

RT performance

RT measures were entered into the same 2x2x2 (Delay type x Congruency x List) ANOVA as above, and the products of this analysis can be found in Table 29. This analysis yielded a no main effect of Delay type, *F*(1,30)=1.40, *p*=.316, List, *F*(1,30)=1.34, *p*=.256, or Congruency, *F*<1. A marginally significant interaction was present between Delay type and List *F*(1,30)=3.42, *p*=.074. This interaction, although only a trend, is in accord with the equivalent interaction noted for accuracy above: Immediate A-B list RTs were slower (*M*=1288 msec, SE=67) than the A-C responses (*M*=1234 msec, *SE*=59) suggestive of RI effects, while in the Delayed condition RTs were slightly faster and closely matched between the A-B (*M*=1165 msec, *SE*=67) and A-C lists (*M*=1178, *SE*=59) suggesting a reduction of RI effects. No interactions were found between Delay type and Congruency, *F*<1, Congruency and List, *F*(1,30)=2.5, *p*=.124, or three-way interaction between these three variables, *F*<1.

Table 29. Experiment 4: Reaction time differences between training and testing across both congruency and list conditions.

	Immediate		Delayed	
	Congruent	Incongruent	Congruent	Incongruent
A-B List	1270 (69)	1306 (72)	1149 (69)	1181 (72)
A-C List	1248 (61)	1221 (64)	1189 (61)	1167 (64)

Note: Values in parentheses indicate the standard error.

Overall these results replicate the observations made in Experiment 3: while levels of A-B performance are maintained across a 24 hour period accuracy levels for the interfering A-C list simultaneously diminish. Furthermore, A-B list accuracy levels were above chance in the Incongruent condition, indicating that despite the large, and negative, impact of incongruent training on memory performance, recovery from RI can now be more definitively said to extend to incongruent information. Lastly, unlike Experiment 3, RTs for list learning measures were also beginning to conform to the reduction of RI effects across time, with a tendency for participants to manifest more equivalent RTs across lists after an extended delay.

4.7.4 2AFC generalisations and novel word size focus

As with Experiment 3, the generalisation measure was split into Narrow (trained size distances ranging from .6 to 2.39) or Wide (trained size distances ranging from 2.65 to 6.15) categories of trained size focus. Responses for these items were assessed together for Congruent and Incongruent generalisations as outlined originally in section 4.3.4.

4.7.4.1 Generalisation % analysis

The factor of Size focus was entered into a 2x2x2 mixed-effects ANOVA featuring the between-participants variable of Delay type (Immediate vs. Delayed) and the between-participants variables of Congruency (Congruent vs. Incongruent) and Size focus (Narrow vs. Wide) using the dependent variable of generalisation %. The products of this analysis can be found in Table 30.

A marginal main effect of Delay type was found, *F*(1,28)=3.58, *p*=.069, with a tendency for higher levels of generalisation % in the Immediate (*M*=66%, *SE*=1.54) relative to the Delayed condition (*M*=62%, *SE*=1.54). As with Experiment 3, a main effect of Congruency was found, *F*(1,28)=5.27, *p*=.029, however in this case it was driven by higher levels of generalisation in the Congruent condition (*M*=66%, *SE*=1.37) relative to the Incongruent condition (*M*=61%, *SE*=1.71). A third main effect of Size focus was present, *F*(1,28)=6.86, *p*=.014, indicating that both conditions were producing higher generalisation scores for novel words with a Narrow (*M*=66%, *SE*=1.22) as opposed to Wide trained Size focus (*M*=62%, *SE*=1.36). No interactions were found between Delay type and Size focus, *F*(1,28)=1.74, *p*=.191, Delay type and Congruency, *F*(1,28)=1.11, *p*=.301, Congruency and Size focus, *F*<1, or three-way interaction between the above factors, *F*<1.

Table 30. Experiment 4: Generalisation performance across 24 hours.

		Immediate		Delayed	
Congruency	Size focus	Generalisation (%)	RTs (msec)	Generalisation (%)	RTs (msec)
Congruent	Narrow	71 (2)	1182 (52)	68 (1.2)	1041 (52)
	Wide	64 (2.6)	1206 (56)	63 (2.6)	1040 (56)
Incongruent	Narrow	68 (2.4)	1201 (55)	57 (2.4)	1044 (55)
	Wide	63 (3.2)	1170 (53)	59 (3.2)	1036 (53)

Note: Narrow and Wide focus items are based on a median split of the rated size distance between two real words (see text). Values in parentheses represent the standard error.

To supplement the above ANOVA the factor of Size focus was broken down into its component units of size distance (ranging from .06 to 6.15) which were then independently correlated with generalisation performance for the Immediate and Delayed conditions. For Congruent generalisations (the top half of Figure 29) these analyses indicated a negative correlation between Size focus and accuracy levels in the Immediate condition, *r*(160)=-.260, *p*=.001, while performance was relatively homogenous across Size focus units in the Delayed condition, although a non-significant negative trend in the same direction was observed, *r*(160)=-.100, *p*=.206. Unlike Experiment 3, a Fisher's r-to-z transformation indicated that the magnitude of this relationship was not significantly different between the Immediate and Delayed conditions Z=-1.47, *p*=.142. For Incongruent generalisations (the bottom half of Figure 29) a marginally significant negative trend was seen in the Immediate condition, *r*(160)=-.142, *p*=.074, while generalisations in the Delayed condition, again, did not display a relationship between Size focus and generalisation %, *r*(160)=-.013, *p*=.872. As with Congruent responses, the magnitude of the Immediate correlation was not significantly different from than that found for Delayed responses, Z=-1.15, *p*=.250.

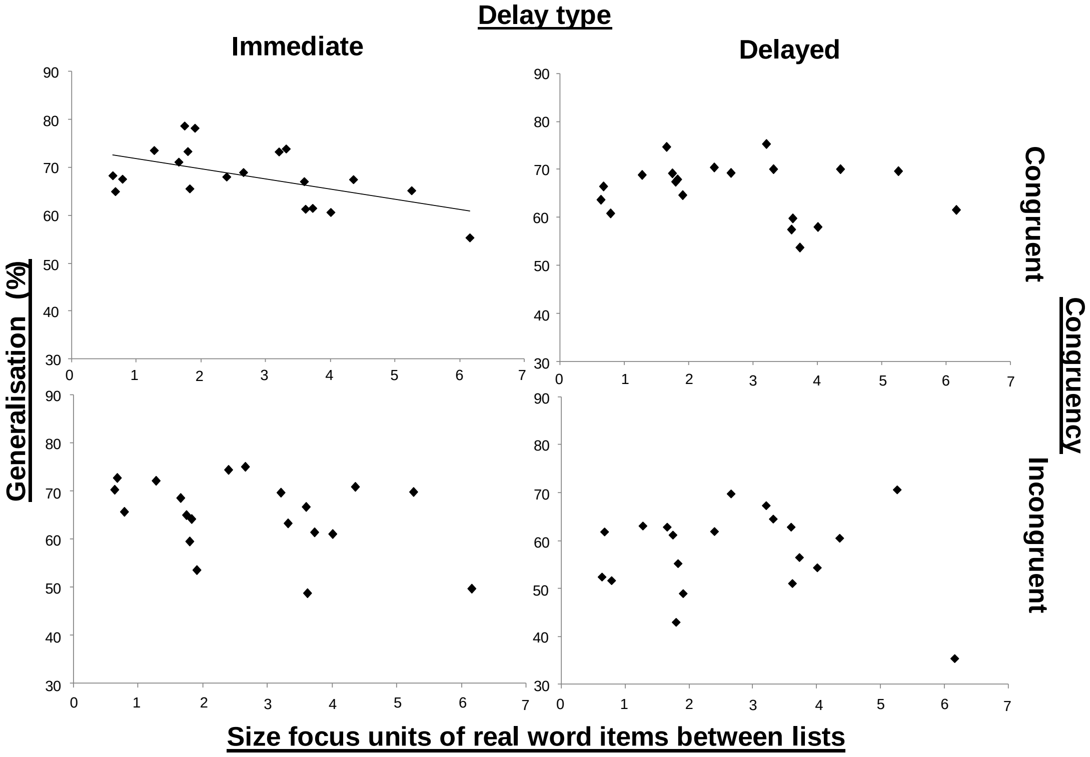


Figure 29. Experiment 4: Generalisation levels (%) correlated with the differences in size focus between real word items for both congruent and incongruent responses across the 24 hour time delay. Note that generalisation scores in this figure are averaged across each size focus unit.

4.7.4.1 Generalisation RT analysis

RT measures were entered into the same 2x2x2 mixed-effects ANOVA as above, and the results of this analysis can also be found in Table 30. Here a marginally significant main effect of Delay type was found, *F*(1,30)=4.03, *p*=.054, with a tendency for globally faster responses in the Delayed condition (*M*=1040, *SE*=53) relative to the Immediate condition (*M*=1040, *SE*=53). No main effect of Congruency, *F*<1, or Size focus, *F*<1 was forthcoming. Moreover, no interactions were found between Delay type and Size focus, *F*<1, or Size focus and Congruency, *F*<1. An interaction was found between Congruency and Size focus, *F*(1,30)=3.21, *p*=.028. Inspection of the data in Table 30 indicates that in the Incongruent condition responses were slower for Wider focus generalisations (*M*=1124, *SE*=40) relative to Narrow focus generalisations (*M*=1112, *SE*=37), in a direction we might expect if Narrow size judgements allowed for more efficient generalisations. However, in the Incongruent condition the above pattern was reversed, with a larger difference between Narrow (*M*=1123, *SE*=39) and Wide focus judgements (*M*=1103, *SE*=1103).

Of the three integrative measures featured in Experiment 4, encoding depth appears to most prominently alter the time related changes in generalisation performance. Here, the impact of size focus on congruent generalisations is now evident immediately, and is not uniquely expressed after 24 hours. A tentative interpretation at this juncture is that the influences of consolidation have been reduced at the higher training threshold featured in this experiment.

With regards to incongruent size generalisations the findings of Experiment 3 were replicated. There was a minimal impact of size focus on the way generalisations were formed over time, although both time delays produced above chance levels of performance. This latter observation indicates that participants remain capable of forming incongruent generalisations within the upper and lower size limits of items presented in training. Furthermore, a new finding specific to this experiment was that RTs displayed a tendency to be globally faster for all generalisation judgements after a 24 hour period, suggesting that some form of consolidation-associated integration may still be emergent on this measure.

4.7.5 Sleep measures

As with the previous experiment there was a large dropout rate of overnight recordings in the Delayed condition due to signal disruption (n=7, 44%). The parameters derived from the remaining 9 intact recordings are displayed in Table 31. Participants slept for an average of 7.2 hours, with a mean sleep onset latency of 16 minutes. These measures reflect the sleep composition expected from a sample of young healthy adults (Spriggs, 2008, p. 15). Indeed comparison with a large corpus of sleep data (Ohayon et al., 2004) corroborates approximate levels of age-appropriate % sleep for TST (~450 mins) and 'Deep sleep' (20%) although 'Light sleep' (~60%) and REM sleep (~22%) may be slightly underestimated.

Table 31. Experiment 4: Sleep parameters.

Sleep parameter	Mean time (min) ± SEM	% of total sleep time ± SEM
Total sleep time	431 ± 14	
Sleep latency	16 ± 3	4 ± 1
Nocturnal wakefulness	6 ± 4	1 ± 1
REM sleep	123 ± 13	28 ± 2
Light sleep (Stages 1 and 2)	223 ± 10	53 ± 3
Deep sleep (Stages 3 and 4)	86 ± 9	20 ± 2

Note: Sleep parameters in parentheses represent the relative sleep stages as defined by Rechtschaffen and Kales (1968), REM = rapid eye movement, SEM = standard error of the mean.

For correlational analysis the same four measures of sleep featured in Experiment 3 were entered into analysis (TST minutes, REM sleep %, 'light' sleep %, and SWS %). These measures were correlated with the same 12 behavioural factors featured in the equivalent analysis conducted in Experiment 3 (please refer to section 4.3.5). No relationships were forthcoming in this analysis, and corresponding table results has been deferred to Appendix 13. Given the large degree of participant drop out it would appear that the power levels have been significantly reduced in these analyses (see Appendix 21).

4.8 Experiment 4 summary and discussion

The aim of Experiment 4 was to explore how higher levels of training (≥95%), relative to the lower levels of training in Experiment 3 (≥80%), might impact the key behavioural findings from the previous investigation. Broadly speaking the majority of the observations made in Experiment 3 were replicated at this higher training threshold. Firstly, there was evidence for the diminished context-specific recall of novel words in the delayed condition, with lower levels of correct list recall concurrent with higher levels of incorrect list recall after 24 hours. Secondly, accuracy performance for the list size judgements indicated RI reduction across both congruent and incongruent word triplets after an extended delay. Moreover, unlike Experiment 3, memory performance was at above chance levels for incongruent A-B list items after a delay, suggesting that RI recovery may be present for both congruent and incongruent materials after 24 hours.

However, in Experiment 4 the impact of size focus on generalisation (%) levels was an effect now expressed 20 minutes after training, and was no longer evident after 24 hours, as in Experiment 3. While this result suggests that consolidation may have less of an influence on the generalisations measures at the higher training threshold in this experiment, generalisations were also accompanied by a tendency to produce faster RTs across 24 hours, indicating that consolidation-associated behavioural changes may still be evident on this measure.

None of the above findings are directly accountable through circadian influences as demonstrated, in particular, by equivalent levels of performance on the newly introduced RT control task. Lastly, there was no evidence of any relationship between the above behavioural changes and the composition of sleep. However, unfortunately, the large dropout rates due to sleep record fragmentation detrimentally impacted the quality of this analysis.

Cued recall measures

The reduction in context-specific recall over 24 hours was replicated in Experiment 4, as participants in the delayed condition recalled fewer items in the correct list context. However, the increased training in this condition also appears to have diminished the number of intrusions on this measure, a factor that was argued in Experiment 3 to potentially reflect the thematic abstraction of semantic materials over time (Diekelmann et al., 2010; Payne et al., 2009). Interestingly, however, there was a corresponding increase in the number of interlist errors after 24 hours not observed in the previous investigation. This latter finding relates to the original predictions made at the start of this chapter, in that 'B' and 'C' list items may be rendered more interchangeable after a consolidation opportunity, thus increasing the number of interlist errors after an extended delay (Inostroza & Born, 2013).

While this finding can still be considered a potential manifestation of decontextualised item memory it should also be noted that the increase in interlist errors over 24 hours was not differentiated by congruency. If this effect was driven by active consolidation processes we might expect error rates to be more pronounced for congruent information as per the original prediction. Moreover, the absolute levels of word recall were diminished in the delayed condition suggesting that the higher levels of training in Experiment 4 potentially reduced the influence of consolidation on this particular measure (Diekelmann et al., 2009; Stickgold, 2009). As with Experiment 3, it remains difficult to distinguish an active role of consolidation in these findings from a passive forgetting of context-specific memory over time (Ellenbogen, Hulbert, et al., 2006). An association between sleep and these shifts in behavioural performance would provide stronger evidence for the former account (Payne et al., 2009) and the appropriate experiment is conducted in the following chapter in lieu of less decisive sleep measures obtained here.

A-B, A-C list size judgements

The selective maintenance of A-B list size relationships across the time delay in the previous investigation was replicated in Experiment 4. Specifically, despite the large differential impact of congruency, a 24 hour time delay was marked by a reduction in A-C list accuracy, accompanied by the relative preservation of A-B list memories. The net result of this effect was that RI effects were removed from both congruent and incongruent size judgements after a consolidation opportunity. Interestingly, if anything, RI recovery was more pronounced at the higher training threshold featured in this investigation. This was noted through a non-significant trend for more accurate A-B list responses relative to the A-C list items after a delay, as well as patterns of RT performance reflective of RI reduction during this time.

Furthermore, unlike Experiment 3, incongruent A-B list performance was now preserved at above chance levels in the delayed condition, while the A-C list accuracy correspondingly dropped to chance levels. This finding indicates two things. Firstly, that the preservation of A-B list associations can be more certainly said extend to incongruent materials, a finding that was not entirely evident from the results of Experiment 3 due the absolute reductions in accuracy for the latter items. Secondly, this finding also suggests that the recovery from RI in the list learning measures may not just derive from a drop in A-C list performance, but also a potential resurgence the memory of the primary A-B list associations (Drosopoulos, Schulze, et al., 2007). This eventuality is discussed in more detail in section 4.9.3 and is explored directly in Experiment 5.

Novel word size generalisations

Of the three behavioural measures present in this experiment the size generalisation measure appears to be most influenced by the higher training threshold. Here, congruent size generalisations reflected the usefulness of novel word size training after 20 minutes, an effect that did not extend across the 24 hour delay. Although the differing impact of B-A-C size training was not statistically distinguishable between the immediate and delayed conditions, the present findings nevertheless suggest that the higher levels of training in Experiment 4 may have reduced the influence of consolidation on this measure (Diekelmann et al., 2009; Stickgold, 2009). That said, there was a tendency for faster responding for all size generalisations after a delay suggesting that some form of ancillary integration may be present on these measures. Indeed, it is interesting to note that this was the only measure in Experiment 4 to show a tendency for faster RTs after a consolidation-opportunity.

In the introduction to this chapter it was suggested that the trained size focus of novel words may impact generalisation performance due the consolidation-associated refinement of information over time. This prediction derives from observations made in the existing literature (Diekelmann et al., 2013; Stickgold & Walker, 2013; Wilhelm et al., 2011) and was provisionally supported by the findings made in Experiment 3. If this selective component of consolidation is mediated by memory strength then integration prior to the offline delay in Experiment 4 may have been sufficient to exceed a level of retention sensitive to the effects of consolidation (Stickgold, 2009).

4.9 Chapter 4 summary and discussion

To date the literature indicates that not all memories appear to benefit equally from sleep and memory consolidation. Instead, consolidation may be particularly attuned to help recover and integrate memories retained at sub-optimal levels (Stickgold, 2009). There is a growing body of evidence that sleep-associated consolidation can actively recover memories from both the non-specific effects of interference accrued across wakefulness (Fenn et al., 2003; Mednick et al., 2002; Wixted, 2004) and more specific forms of interference induced by memories with conflicting, or shared, components (Abel & Bäuml, 2013; Drosopoulos, Schulze, et al., 2007; Ekstrand, 1967). Although the recovery of memories from the influences of RI can be considered adaptive (Fenn et al., 2003) it was noted in the introduction to this chapter that there must be a selective component to RI recovery to circumvent the influx of maladaptive information into long-term memory (Stickgold & Walker, 2013).

To explore how the usefulness of recently acquired memories may influence RI recovery and integration the two experiments in this chapter featured a modified A-B, A-C list learning paradigm. Here, changes to memory across a 24 hour period were examined in relation to word lists that were semantically compatible (congruent) or incompatible (incongruent) with prior knowledge. Three measures were implemented: (i) a traditional measure of list memory (cued recall), (ii) a measure of memory relating to the specific size relationships established across lists (2AFC list judgements), and (iii) a measure designed to probe the integration of novel words in relation to new, but familiar information (2AFC size generalisations). The depth of encoding was also manipulated between experiments 3 and 4 to further explore how differing levels memory strength may influence any temporal changes to behaviour.

The key findings from these experiments can be summarised as follows. Firstly, RI and congruency effects were not observed on the cued recall measures, although there was evidence for reduced context-specific list memory after 24 hours in both experiments. Secondly, in contrast to the cued recall measures, RI effects were present for memories relating to the size relationships established across the A-B and A-C word lists 20 minutes after training. Moreover, these RI effects dissipated over 24 hours, suggesting that consolidation processes may be facilitating RI recovery. Thirdly, the reduction of RI effects over time was also associated with an absolute reduction in the capacity to remember incongruent size relationships in both experiments, suggesting a selective component to RI recovery. This latter effect first emerged 20 minutes after training, and became more pronounced after 24 hours, suggesting that incongruent list memories were rapidly forgotten relative to congruent condition.

With regards to encoding depth, a lower training threshold in Experiment 3 resulted in size generalisations that reflected the usefulness (i.e. size focus) of size relationships established across the B-A-C word triplets after 24 hours. However, a higher training threshold in Experiment 4 produced a similar effect in the immediate condition with an extended time delay instead associated with a tendency for faster performances this measure. When taken together these observations suggest that the potential consolidation-associated integration of novel words is contingent on memory strength, as the impact of size focus was more prominently expressed at a lower training threshold.

4.9.1 RI recovery and prior knowledge

At a behavioural level the reduction of RI effects across 24 hours in Experiments 3 and 4 are compatible with the observations made in several classic studies implementing the A-B, A-C list learning paradigm. For example, Underwood (1948a) trained participants on a series of semantically unrelated adjectives using A-B, A-C list learning protocols with follow up testing administered either 5 hours, or 48 hours, after training. Here, it was found that the influences of RI seen after 5 hours had diminished at the later time point, as expressed through a relative drop in A-C list recall. In a similar investigation the time delay required to dissipate RI effects was found to occur within 24 hours, a time frame that presumably involved a period of sleep (Underwood, 1948b). Indeed, the temporal nature of RI reduction has been replicated in several other investigations including variations of these procedures featuring interpolated lists containing nonsense syllables (Ceraso & Henderson, 1965) and letters of the alphabet presented in strings of Morse code (Rothkopf, 1957).

However, unlike these previous studies, a novel aspect of Experiments 3 and 4 was that RI effects, and their subsequent time-dependent dissipation, were observed exclusively for memories relating to the size relationships established across the A-B and A-C word lists. Conversely, the equivalent effects were not found on the more traditional measures of cued recall (Barnes & Underwood, 1959). Previous studies exploring the influence of semantic relationships on A-B and A-C list learning have not typically probed the influence of time memory performance. What evidence there is indicates that RI effects can indeed emerge shortly after training for list items that share semantically themed adjectives (e.g. ‘*weep-cry*’ and ‘*weep-sob*’) (Baddeley & Dale, 1966) or nouns (historical figures or geographical locations) (Bower et al., 1994). The present results therefore extend these observations by showing that RI effects can also be ameliorated for the semantic information formed across B-A-C word triplets after a sufficient time delay.

4.9.2 Theoretical accounts of RI

Over 70 years of research into the phenomenon of interference and forgetting has produced a wealth of theoretical accounts on the subject (Anderson, 2003; McGeoch & Irion, 1952; Melton & Irwin, 1940; Postman & Stark, 1969; Wixted, 2004). Although the experiments presented in this chapter were not designed to differentiate between these accounts they nevertheless help expand on the conceptualisation of the interference phenomenon. The present findings will therefore be briefly discussed in relation to three relevant models of interference: the response-set suppression hypothesis (Postman & Stark, 1969), the executive inhibition account (Anderson, 2003), and the unlearning account of retroactive interference (Melton & Irwin, 1940).

The response-set suppression hypotheses proposes that RI effects will arise due to the diminished availability of *all* A-B list responses once the secondary A-C associations have been acquired (Postman & Stark, 1969). A noted issue with this account is that it deemphasizes the role that interlist relationships may play in modifying the impact of RI on memory (Bower et al., 1994). Indeed, counter to this model, the experiments presented in this chapter demonstrate that interlist relationships alone are sufficient to induce RI. Moreover, the congruency of these relationships had a clear influence on list memory, as reflected in the low levels of accuracy for incongruent size relationships at both experimental time points. Neither of these observations are accommodated by the response-set suppression hypothesis, particularly given that the real word items were counterbalanced equally across the congruency manipulation.

A more recent account of RI argues that interference will arise primarily at retrieval through the active suppression of A-B associations through executive processes (Anderson, 2003). This model proposes that conflicting memories (i.e. the primary A-B list associations) need to be ‘overridden’ at test to preventing their unwanted intrusion at recall. This account, in contrast to the response-set suppression hypothesis above, places more of an emphasis on how interlist relationships may influence RI. Accordingly, executive inhibition may explain the minimal impact of both RI and the congruency manipulation on the cued recall measures. As cued recall does not probe the size relationships established across the B-A-C word triplets interlist conflict may have been minimised on this particular measure. Conversely, interlist conflict may have been maximised in the incongruent condition, as the real word items in both the ‘B’ and ‘C’ list contained contradictory information regarding the size of the shared ‘A’ item. As such, an active suppression of these memories at test may explain the rapid drop in accuracy in this condition.

While an executive inhibition account explains why RI effects were predominantly expressed for the size relationships established across word lists this model has little to say about temporal nature of the RI recovery observed in Experiments 3 and 4. This time-dependent change to list memory bears particular relevance to Melton and Irwin's (1940) classic unlearning account of retroactive interference. Here it is proposed that RI may occur, in part, through the temporary unlearning of A-B list memories when these associations remain unreinforced during the acquisition of the secondary A-C list. This account draws parallels with the phenomenon of extinction in classical conditioning literature. As extinguished responses often spontaneously recover after a sufficient period of time (Rescorla, 2004) weakened A-B list associations, too, are thought to recover after an extended delay. Unlearning has been repeatedly invoked to explain the temporal recovery of declarative memories from RI in previous experiments (Underwood, 1948a, 1948b) although the precise mechanisms underlying this effect are not fully understood (Postman & Underwood, 1973).

The observation that the memory of A-B list size relationships are, at the very least, selectively maintained across a 24 period in Experiments 3 and 4 provisionally conform with the unlearning account (Melton & Irwin, 1940). Although the spontaneous recovery of word pair associations from RI has been found in to occur within the first 20 minutes after training in some experiments (Ekstrand et al., 1971; Postman, Stark, & Henschel, 1969) recovery in these instances appears to be the product of explicit rehersal during this time (Drosopoulos, Schulze, et al., 2007). Unlike these experiments spontaneous recovery did not occur within a 20 minute timeframe in either of the investigations contained in this chapter. Moreover, in Experiment 4, only incongruent A-B list associations remained at above chance levels after an extended delay, suggesting that it is these particular associations that may benefit the most from a consolidation-opportunity. Although, more data is needed to fully corroborate this account (as will be discussed) the present data remain provisionally compatible with the growing body of research implicating the sleep state in the active recovery of memories from RI (Abel & Bäuml, 2013, 2013; Ekstrand, 1967).

4.9.3 The potential role of sleep in mnemonic recovery and memory integration

Current models of sleep and memory consolidation supplement Melton and Irwin's (1940) unlearning account by outlining a set of testable mechanisms responsible for the reactivation and reprocessing of memory. Indeed, Drosopoulos, et al., (2007) propose that the systems-level interaction between the hippocampus and neocortex during SWS may not only explain spontaneous recovery but also more the more general recuperative benefits of sleep on declarative memory (Fenn et al., 2003; Mednick et al., 2002; Schmidt et al., 2006). In line with this account recent fMRI evidence exploring the neurological correlates A-B and A-C list learning has revealed a strong association between hippocampal activity and the resistance of A-B list memories to the detrimental influences of RI (Kuhl, Shah, DuBrow, & Wagner, 2010). As sleep-associated consolidation has been causally linked to latent reactivation of hippocampally-dependent memories (Rudoy et al., 2009; Wilson & McNaughton, 1994) the notion that this state may facilitate the recovery of A-B list memories from RI is not unreasonable.

As noted in the discussions for Experiments 3 and 4 the present findings remain compatible with the existing RI and consolidation literature. However, it is important to consider alternative explanations for these results. In Chapter 1 (section 1.3.2) an ongoing debate was introduced regarding the active or passive protective influences of sleep on memory performance (Ellenbogen, Payne, et al., 2006). Specifically, an active role of sleep is sometimes difficult to distinguish from the passive-protection from waking interference that this state may provide (Wixted, 2004). As the effects of time were not controlled for in either of the experiments presented in this chapter the present findings may be the product of memory decay across wakefulness. It should be noted that due to the (deliberately) variable times of training and testing in the delayed condition (9am-6pm) some participants may have been awake for up to 16 hours prior to a sleep opportunity, leaving ample time for waking interference to potentially influence memory.

Relevant to this notion Slamecka and Ceraso (1960) propose that the dissipation of RI effects across time may the product of a differing vulnerability of A-B and A-C list memories to forgetting. Specifically, the acquisition of the A-C list may retroactively suppress, or remove, the weakest associations in the primary A-B list (McGeoch & Irion, 1952; Postman & Underwood, 1973). Conversely, as A-C items are not influenced by RI, the secondary list may contain a greater number of relatively weak associations. Accordingly, two separate forgetting curves may apply to the A-B and A-C list associations, a slower rate of decay for A-B list memories and a faster rate of decay for A-C list memories. All things being equal, the passage of time may induce a larger drop in the memory performance for A-C associations, eventually leading to equivalent levels of recall between the two lists. Although previous investigations exploring the specific role of sleep in RI recovery raise doubts about this simple forgetting account (Abel & Bäuml, 2013; Drosopoulos, Windau, et al., 2007; Ekstrand, 1967) this explanation nevertheless cannot be fully discounted regarding the present results.

A similar account can also be applied to the changes seen in size generalisation performance over 24 hours. For example, novel words with a narrow size focus (e.g. rat>shusk>flea) may be less prone to waking interference than novel words with a wider degree of inferred size (e.g. horse>snafe>rabbit). Due to the relative differences in specificity regarding the trained size of novel words narrow and wide focus items may produce stronger and weaker representations in memory respectively. As such, generalisation performance for the latter items may be more vulnerable to waking decay (Lechner et al., 1999; Wixted, 2004). One argument against this interpretation is the observation that a higher training threshold in Experiment 4 induced the impact of size focus just 20 minutes after training. However, as with the primary list learning measures, this account cannot be dismissed without further evidence.

Lastly, the effects of forgetting and waking interference also extend to the cued recall measure. As noted in the discussion of Experiment 3, the increased number of untrained intrusions after 24 hours may be reflective of increased guesswork on the part of the participants due to poorer item memory. Similarly, the diminished context-specific recall after 24 hours in both Experiments 3 and 4 may be attributed to the passive decay of contextual information across wakefulness (Wixted, 2004). Here, as with all of the alternative accounts outlined in this section, an association between these measures and sleep physiology would provide stronger evidence for an active role of sleep in memory consolidation (Ellenbogen, Payne, et al., 2006).

4.9.5 Conclusions

Taken together the findings of Experiments 3 and 4 suggest that the changes to memory that can arise over 24 hours may confer several benefits to future behaviour. While useful (congruent) size relationships can recover from the short-term influences of RI, memories for less useful (incongruent) size relationships are simultaneously diminished over same time period. Moreover, in Experiment 3 the generalisation measure indicated that the usefulness of the size relationships established across the B-A-C word triplets will only influence performance after 24 hours. As this effect arose immediately at higher training threshold in Experiment 4, there is evidence that selective integration on this measure may only be present for memories retained at a lower threshold.

Although all three behavioural measures in this chapter revealed interesting changes to memory across 24 hours period it remains difficult to fully distinguish the relative contributions of sleep and wakefulness on performance, particularly given the inconclusive sleep measures featured in both investigations. The case for an active role for memory consolidation would therefore be strengthened by (i) a meaningful relationship between sleep physiology and memory performance, and (ii) a systematic exploration the differential impact of sleep and wakefulness on the time-dependent changes to memory uncovered in this chapter. Experiment 5, presented in the following chapter, explores these issues directly.

Chapter 5: The role of sleep and wakefulness on mnemonic recovery and integration.

5.1 Introduction

Using a modified A-B, A-C list learning paradigm, the findings from Experiments 3 and 4 indicate that consolidation may have a role in both the recovery and integration of recently acquired memories. Firstly, a time delay of 24 hours, relative to 20 minutes, was associated with diminished retroactive interference (RI) effects for size relationships established across the A-B and A-C word lists. Secondly, the compatibility of these relationships with prior knowledge appeared to influence this recovery, with memory of incongruent associations otherwise reduced to chance, or near chance, levels after an extended delay. Thirdly, in terms of integration, Experiment 3 found that the usefulness (i.e., size focus) of real word relationships across the B>A>C word triplets influenced the formation of congruent, but not incongruent, novel word generalisations after 24 hours. Together these key findings implicate a degree of selectivity in the way memories both recover and integrate over time, particularly for information trained at a lower threshold.

As discussed in Chapter 4 the observations made in these previous experiments mesh well with the findings from several classic studies exploring the temporal nature of RI recovery (Rothkopf, 1957; Underwood, 1948a, 1948b) and more recent investigations relating to the role that consolidation may have in this regard (Abel & Bäuml, 2013; Drosopoulos, Schulze, et al., 2007; Ekstrand, 1967). RI recovery in this manner may also be considered adaptive, as sleep may help remove the influences of RI accrued across the waking day (Fenn et al., 2003). However, it was also noted that the inconclusive sleep measures in Chapter 4, in conjunction with the use of immediate and delayed experimental time points, limited the number of conclusions that could be drawn from these measures. Two principle questions were raised as a result: (i) to what extent are some of these behavioural changes observed over time attributable to active consolidation processes or differential rates of forgetting?, and (ii) what mechanisms, if any, may underlie RI recovery?

To help resolve these questions Experiment 5 applies the modified A-B, A-C paradigm featured in Experiment 3 to probe memory at two additional time points: a 12 hour period involving either sleep or wakefulness as part of a standard 12:12 sleep/wake design. Accordingly, this chapter will provide important information regarding the differential impact of sleep and wakefulness on changes to memory that occur within a 24 hour time period. Moreover, high-quality Polysomnographic (PSG) sleep measures replace the less robust sleep measures featured in Chapter 4. Experiment 5 will therefore provide a more fine-grained set of analyses regarding the potential relationship between sleep physiology, RI recovery, prior knowledge and memory integration.

5.1.1 Potential consolidation-associated mechanisms of RI recovery and integration

In Chapter 4 an account of RI was introduced which attributes interference effects, in part, to the temporary ‘unlearning’ of redundant associations (Melton & Irwin, 1940). In terms of the A-B, A-C paradigm, memories of the primary A-B list are thought to be suppressed through their lack of reinforcement during the acquisition of the superseding A-C list. This account considers unlearning comparable to the phenomenon of extinction in the classical conditioning literature, where previously suppressed memories are often found to spontaneously recover after a sufficient period of time (Rescorla, 2004). Similarly, spontaneous recovery from RI is considered to be a time-dependent phenomenon which can prompt the eventual resurgence of weakened A-B list, but not A-C list, memories. Although the parallels with classical conditioning have proved to be an enduring and useful analogy for spontaneous recovery (Postman & Underwood, 1973) the precise mechanisms underlying this effect have, up until recently, remained rather vague.

To this end, memory consolidation supplements Melton and Irwin's (1940) unlearning account by providing a set of endogenous processes responsible for the post-encoding reactivation and reprocessing of memories (Frankland & Bontempi, 2005; McClelland et al., 1995). Drosopoulos et al. (2007) suggest that spontaneous recovery may be particularly contingent on the interaction between hippocampal and neocortical memory systems during the sleep state. Current models of consolidation indicate that the ongoing acquisition of new memories across wakefulness will involve the joint activation of both prefrontal and hippocampal networks (Dudai, 2004; Sterpenich et al., 2007; Takashima et al., 2006). However, the reprocessing of hippocampally held memories during slow wave sleep (SWS) is thought to occur at a point when prefrontal activity is relatively diminished (Maquet, 2000). Subsequently, the signal output from the hippocampus to the neocortex during these stages may not be differentiated between weaker and stronger memories, resulting in the relatively enhanced signal transfer of weakly encoded information. The undifferentiated amplification of memories retained at sub-optimal levels during SWS may explain the general recuperative benefits of consolidation on memory (Diekelmann et al., 2009; Stickgold, 2009) as well as the specific recovery of memories temporarily weakened by interference (Abel & Bäuml, 2013; Drosopoulos, Schulze, et al., 2007).

Furthermore, with regards to the influences of prior knowledge and word triplet congruency, there is reason to consider memories that share overlapping contextual features as being particularly amenable to consolidation-associated recovery. A recent fMRI study exploring the neural correlates of A-B, A-C list retention found that hippocampal activation during the encoding of specific B-A-C word triplets was predictive of the subsequent resistance of A-B list memories to the detrimental influences of RI (Kuhl et al., 2010). Moreover, the shared context across lists during learning, which was manipulated by priming associations with financial incentives of varying strength, was found to produce a strong co-activation of reward-associated regions (the ventral striatum) and the hippocampus at recall. The co-activation of these areas was in turn predictive of superior memory retention of the B-A-C word triplets. Although this particular investigation only probed memory retention directly after training these findings are nevertheless compatible with the notion that consolidation may confer additional benefits to mnemonic recovery.

Firstly, hippocampal activation during encoding is considered a pre-requisite for subsequent systems-level processing in the standard model of memory consolidation (Frankland & Bontempi, 2005, 2006). Indeed, the overlapping semantic and associative features shared across congruent word triplets (e.g. 'B>A>C') may strongly prime the hippocampal system in this regard (Heckers et al., 2004; Lewis & Durrant, 2011). Secondly, the shared size context across congruent word triplets may provide mutually reinforcing cues should these memories become reactivated during consolidation (Kuhl et al., 2010). Thirdly, the presence of relevant elements of prior knowledge (the B and C list items) may bolster this process further by facilitating the rapid assimilation of these shared representations into pre-existing memory (Tse et al., 2007; van Kesteren et al., 2012). All three of these factors may contribute to a particularly strong resurgence in the memory of 'unlearned' A-B list associations after a period of consolidation.

The levelling of output signals from the hippocampus during SWS (as noted above) may also facilitate the discovery of shared representations during consolidation (McClelland et al., 1995). The observation in Experiment 3 that a period of time is required for generalisations to reflect the inherent usefulness of word triplet relationships is therefore compatible with the notion that systems-level processes can help assimilate newly acquired episodic memories into pre-existing semantic networks (Rogers & McClelland, 2004). Additionally, this finding fits well with the broader literature suggesting that the consolidation of new information is often contingent on its applicability to future behaviour (Diekelmann et al., 2013; Stickgold & Walker, 2013; Wilhelm et al., 2011).

Lastly, although the cued recall measures featured in Experiments 3 and 4 did not appear to be influenced by either RI or congruency, both investigations found evidence for diminished context-specific recall over time. This was seen after 24 hours in both experiments alongside a tendency to recall untrained intrusions after an extended delay in Experiment 3. As noted previously these changes in behaviour may reflect the progressive decontextualisation of episodic information, an effect known to be facilitated by consolidation (Inostroza & Born, 2013). Indeed, these findings are provisionally compatible with the false memory literature showing that consolidation may enhance the thematic 'gist' of information while simultaneously reducing the need to retain episodic details (Diekelmann et al., 2010; Payne et al., 2009). It was interesting to note that a time delay in Experiment 3 did not influence the absolute levels of recall, yet a larger number of untrained intrusions were present after 24 hours. This qualitative shift in the expression of memory over time may reflect an increased capacity for generalisation, a purported advantage of gist abstraction (Reyna & Kiernan, 1994).

5.1.2 Alternative accounts

Although the results from the previous two experiments are readily compatible with the literature relating to memory consolidation several issues were raised at the end of Chapter 4 that caution against an overinterpretation of these data. Firstly, the predicted associations between the changes in memory and sleep physiology were mostly absent. Although there was the suggestion that Rapid Eye Movement (REM) sleep may be involved in the recovery of congruent A-B list memory retained at a lower threshold in Experiment 3 (Norman et al., 2005) the high levels of participant drop-out using the Zeo sleep-recording kits rendered these measures largely inconclusive.

A second issue relates directly to the time delay manipulation featured in Experiments 3 and 4. While the use of 20 minute and 24 hour time points provided insights into how the passage of time may influence memory, the relative contributions of sleep and wakefulness on these outcomes were not dissociated. Relevant to this issue is the notion that wakefulness contributes to the progressive forgetting of information. As we constantly acquire new information in our daily lives memories formed earlier in time may become increasingly vulnerable to the non-specific effects of waking interference (Wixted, 2004). Moreover, in this instance, the benefits of the sleep on memory may be considered passive-protective, with the relative neurobiological quiescence of the sleep state providing respite from the progressive memory loss induced by wakefulness (Ellenbogen, Payne, et al., 2006; Payne et al., 2012). With regards to the key findings made in Experiments 3 and 4 it remains difficult to distinguish this account from the potentially active role of consolidation.

5.1.3 Experiment 5 aims and predictions

It should be clear from this introduction that for an active account of consolidation to be fully corroborated a systematic manipulation of sleep and wakefulness is required. To do this Experiment 5 extends the paradigm featured in Experiment 3 to include two further time points involving a 12 hour period of nocturnal sleep or wakefulness. The decision to use the lower training threshold featured in Experiment 3 (≥80%) and not the higher training threshold in Experiment 4 (≥95%) was made on the basis of the literature reviewed in Chapter 4 (section 4.1) which suggests that memories retained at suboptimal thresholds are generally more amenable to the influences of consolidation (Diekelmann et al., 2009; Stickgold, 2009). Indeed, this contention is in line with at least two observations made in Experiment 3: (i) the impact of the size focus factor on the congruent generalisation measures after 24 hours, and (ii) the similar levels of absolute recall (regardless of list or congruency) between the immediate and delayed conditions, which were otherwise found to diminish over time at the higher training threshold featured in Experiment 4.

5.1.3.1 Behavioural predictions

For cued recall measures the active and passive accounts outlined in the previous sections form distinct predictions with regards to the manipulation of sleep and wakefulness. An active account would implicate systems-level consolidation in reduction of context-specific memory (Inostroza & Born, 2013) thus promoting increasingly interchangeable size relationships over the B-A-C word triplets after sleep. Moreover, should this effect occur in conjunction with extraction of thematic gist (i.e. animals), we might also expect a larger number of untrained intrusions on the recall measures in this condition (Diekelmann et al., 2010; Payne et al., 2009). Conversely, should these effects reflect diminished context-specific memory as a product of waking interference then the equivalent findings should emerge after a period of wakefulness. Moreover, in this latter scenario, sleep may help protect memories from waking interference, thus conferring a lower number of inter-list errors and intrusions after 12 hours.

The passive vs. active accounts of RI recovery also make divergent predictions regarding the impact of sleep and wakefulness on 2AFC list memory. For example, Slamecka and Ceraso's (1960) account of differential forgetting would predict a greater reduction in A-C list memory over wakefulness, as it is these associations that are deemed to be the most vulnerable to temporal decay. However, the influences of RI should remain relatively preserved after a period of sleep as interference is minimised during this state. On the other hand, should sleep actively facilitate the spontaneous recovery of temporarily weakened A-B list associations, then memory performance on these items should spontaneously recovery in this condition, while RI effects are simultaneously preserved across a period of wakefulness (Drosopoulos, Schulze, et al., 2007). These differing accounts relate specifically to measures of 2AFC list memory, as the influences of RI in Experiments 3 and 4 were restricted to this measure.

An additional note should be made regarding the effects of congruency on the 2AFC list measure. In Chapter 4 the rapid drop in accuracy for the incongruent size relationships after 20 minutes suggests that memory for these associations was influenced by heightened response competition at test (Anderson, 2003) with a potential role for consolidation deferred to the ongoing suppression of these memories (Tononi & Cirelli, 2003, 2006). It remains harder to apply the active vs. passive dichotomy framing this experiment to the incongruous list measures, as both consolidation and waking interference may have a role in the progressive loss of memory for these size relationships. Moreover, poorer memory performance after a period of sleep may also reflect the absence of consolidation processes that are otherwise predicted to support the synthesis of congruent information. To clarify these eventualities measures of sleep physiology are required (outlined in the following section).

Lastly, the impact of trained size focus on generalisation performance confers two relatively straightforward predictions. An active consolidation account would predict the selective enhancement of narrow focus novel word generalisations after a period of sleep. This may be instigated through the bi-directional interaction between episodic and semantic memory systems through consolidation (McClelland et al., 1995) and the selective retention of information with more value to future behaviour (Stickgold & Walker, 2013). On the other hand, if wider focus size judgments are more vulnerable to waking decay, we would expect the impact of size focus to emerge on measures of waking generalisation only. It is not clear if these effects will extend to incongruent novel word generalisations given the undifferentiated impact of the time delay manipulation on these measures in Experiments 3 and 4. However, as incongruent word triplets may result in less stable size inferences we might expect these generalisations to be inherently more vulnerable to waking interference.

5.1.3.2 Predictions relating to sleep physiology

Support for an active account of consolidation will be strengthened through the observation of a meaningful relationship between sleep physiology and one or more of the behavioural effects predicted to arise after a sleep opportunity. As established in Chapters 1 and 3, SWS is considered to be the most optimal state for systems-level consolidation to occur (Diekelmann & Born, 2010b; Steriade, 2001; Stickgold, 2005). Moreover, the literature reviewed here suggests that SWS may be particularly beneficial for weakened memories (Göder & Born, 2012; Maquet, 2000) and memories that share overlapping contextual features (Lewis & Durrant, 2011; McClelland et al., 1995).

Additionally, we may also find that the reduced memory for incongruently trained word lists is facilitated though synaptic downscaling processes also thought to occur during SWS (Tononi & Cirelli, 2003, 2006). Taken together, the present literature is aligned on the prediction that SWS in particular may be associated with the selective recovery, refinement and integration of memory in Experiment 5. This eventuality does not preclude the potential involvement of rapid eye movement (REM) sleep in these behavioural outcomes

(Norman et al., 2005; Walker & Stickgold, 2010) although the empirical grounding for this latter prediction is generally considered less robust (Diekelmann & Born, 2010b).

As with Experiment 2, the present investigation will feature measures of sleep spindle activity given their prominent association with systems-level consolidation and mnemonic integration (Clemens et al., 2007; Mednick et al., 2013; Tamminen et al., 2010). Moreover, Experiment 5 will feature measures of fast (13-15 Hz) and slow (11-13 Hz) spindles that reside within the standard spindle frequency range (11-15 Hz). Fast and slow spindles are considered to be topographically dissociable, with distributions focussed over centro-parietal and anterior regions of the scalp respectively (Zeitlhofer et al., 1997). Fast spindles may have more functional involvement in systems-level consolidation as they have been found to display a closer relationship with hippocampal activity during non-rapid eye movement (NREM) sleep (Schabus et al., 2007) and a more pronounced synchronisation with SWS activity (Andrillon et al., 2011). That said, fast and slow spindles have been associated with declarative memory consolidation both independently (Schmidt et al., 2006; Wilhelm et al., 2011) and in parallel (Schabus et al., 2006; Tamminen et al., 2010) suggesting that a more fine-grained analysis would benefit from examining these spindle types independently.

5.2 Methods

The methodology for this experiment is identical to that used in Experiment 3 (Section 4.2). As such only differences from these protocols are outlined and justified here.

5.2.1 Participants

Thirty-seven participants aged between 18 and 27 were recruited through the University of York Psychology electronic experiment booking system (PEEBS). All were healthy, non-smoking native English speakers with no prior history of sleep or psychiatric disorders. All participants agreed to abstain from caffeine, alcohol and napping over the course of the experiment and 24 hours beforehand. Written informed consent for participation was obtained in compliance with the local ethics committee.

Two participants were removed from analysis due to excessively slow RTs (>2.5 standard deviations) at training (n=1) and at both training and test (n=1). The remaining participants were randomly assigned to the wake (n=17; 11 females, mean age = 19.7, *SD*=2.6) or sleep conditions (n=18; 8 females, mean age = 20.3, *SD*=2.3). Participants had a mean age of 20.3 years (*SD*=2.2) and were provided payment or course credit for taking part, with an additional £20 Amazon gift voucher awarded to the best overall performer.

5.2.2 Design

The present experiment, as seen in Figure 30 (A), features a mixed design containing the between-participant variable of Delay type (Wake or Sleep) and the within-participant variables of Congruency (Congruent or Incongruent), List (A-B list memory vs. A-C list memory) and the trained Size focus of novel words (Narrow vs. Wide).

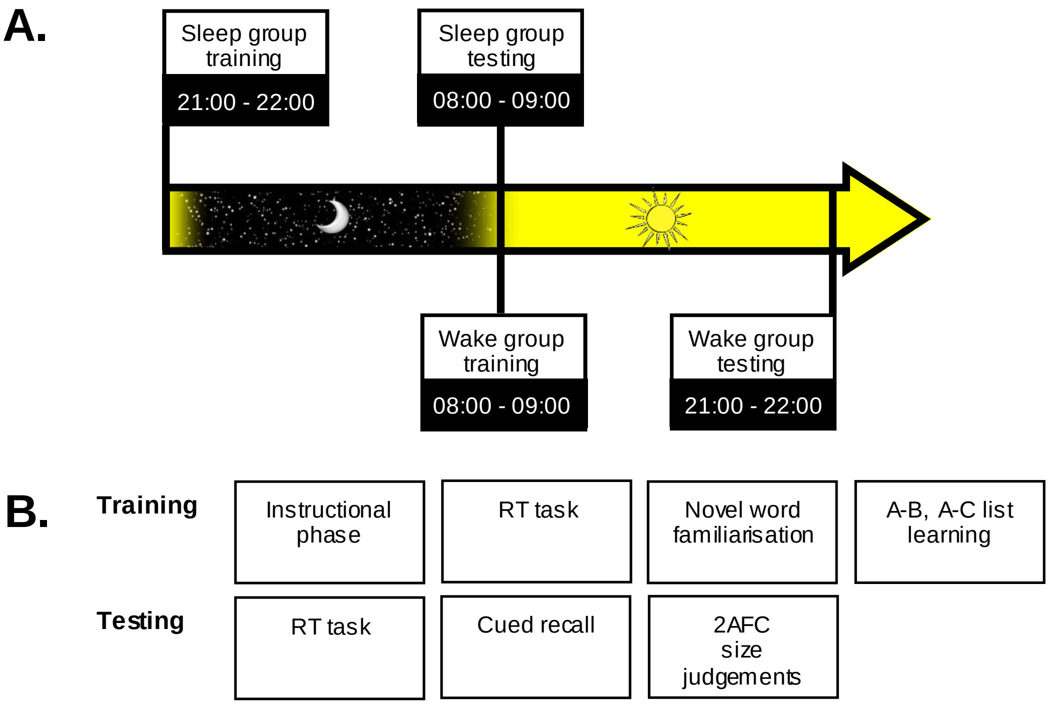


Figure 30. Experiment 5 design. (A) The between-participant conditions in this experiments. (B) The sequential order of tasks across training and testing.

Participants were randomized into sleep and wake conditions following a standard 12:12 sleep/wake design (Clemens et al., 2005; Tamminen et al., 2010; Wilhelm et al., 2011). Participants in the wake condition started the training phase between 9:00 am and 10:00 am in the morning. After training was completed they spent the rest of the waking day performing their regular activities precluding napping and the consumption of caffeine and alcohol. Participants returned to the laboratory 12 hours later to begin the testing phase, which was initiated between 9:00 pm and 10:00 pm in the evening. Conversely, participants in the sleep condition arrived at the laboratory between 9:00 pm and 10:00 pm in the evening after dinner where they underwent preparation for polysomnography (PSG) recordings prior to the training phase. Once training was completed, participants were immediately put to bed between 11 pm and 12 pm and woken up by the experimenter at around 7:45 am (after ~8hrs sleep) with the option of a light snack and a shower. Morning testing occurred between 30 and 40 minutes after arousal to avoid the non-specific cognitive detriments of sleep inertia (Gais et al., 2002; Tamminen, 2010).

5.2.3 Procedure

The stimuli, instructions and experimental tasks are identical to those used in Experiment 3, however the baseline RT control task adapted from Nishida and Walker (2007), used previously in Experiment 4 (see section 4.6.1), was retained for this investigation (Nishida & Walker, 2007). This RT task provides an objective gauge of alertness to supplement the more subjective alertness measures obtained using the Stanford Sleepiness Scale (SSS) in this experiment (Hoddes et al., 1973).

5.2.4 PSG recording and analysis

PSG recordings were obtained using a Grass Technologies system with impedance levels, filter settings and sampling rates calibrated in accordance with the standardized American Academy of Sleep Science Manual (AASM, 2007). Recordings were made digitally using an Embla© N7000 headbox and reusable Biosense© brand gold-disk electrodes. Electrode placement followed the standardized 10-20 scheme (Jasper, 1958) using an eight channel montage with six bilateral scalp electrodes (O1, O2, C3, C4, F3, F4) each referenced to the contralateral mastoid (M1, M2). This montage differs from that used in Experiment 2 in that frontal electrodes are now included. Frontal placement should provide good measures of SWS activity and more accurate measures of slow spindles (11-13 Hz), which are thought to manifest a more frontal topographic distribution (Schabus et al., 2007; Zeitlhofer et al., 1997). Two electro-oculographic (EOG) channels were used to monitor eye movements (LOC, ROC), and two electromyographic (EMG) channels monitored chin movements (X1 and X3). Montage sites were cleaned using NuPrep© exfoliating gel and electrodes were affixed by means of EC2© electrode cream and surgical tape. All electrodes had impedance levels less than 5kOhms with the recording signal digitally sampled at 200Hz.

PSG Scoring

Three sleep records had to be removed from analysis due to an equipment failure on the experimental night. Sleep scoring on the remaining 15 intact sleep records was conducted manually using Remlogic© 1.1 software using the referenced C4-M1 channel in 30 second epochs. Scoring adhered to the guidelines established by the American Academy of Sleep Medicine (AASM) thus providing measures of REM sleep, stage 2 sleep, and SWS (Danker-Hopfe et al., 2009). The AASM scoring guidelines thought to provide a more biologically plausible, and simpler, set of sleep scoring criterion than the related Rechtschaffen and Kales (1968) scoring system used in Experiment 2 (Himanen & Hasan, 2000).

Sleep spindle detection

As with Experiment 2, sleep spindles were detected using an algorithm originally developed by Ferrarelli et al. (2007) which has been validated in several published studies (Nishida & Walker, 2007; Tamminen, 2010). EEG recordings were artefact rejected before being bandpass filtered independently for spindles within the standard (11-15 Hz), fast (13-15 Hz) and slow (11-13 Hz) frequency ranges using a linear finite impulse response filter in EEG lab toolbox for Matlab (http://www.sccn.ucsd.edu/eeglab/). Spindle data was extracted using Matlab (The MathWorks Inc, Natick, MA) based on the number of discrete spindle events for each channel. Amplitude fluctuations in the unfiltered time series exceeding a specific threshold were counted as spindles. This threshold was established as eight times the amplitude in the signal of a given EEG channel, so as to account for individual variations in spindle expression (Ferrarelli et al., 2007).

5.2.5 Other measures

As with all of the experiments presented in this thesis, the Stanford Sleepiness Scale (SSS) was used to assess participant alertness levels (Hoddes et al., 1973). This itemised measure required participants to rate their levels of alertness ranging from 1 (highly alert) to 7 (sleep onset) prior to training and testing. Participants were also asked to report their average sleep time (in hours) across the two weeks prior to the experiment as well as the amount of sleep obtained prior to the testing phase. For the wake condition this latter measure referred to the night of sleep prior to coming into the lab, whereas for the sleep condition this referred to the night of sleep obtained between the training and testing phases. Lastly, participants were also asked to rate the quality of this sleep period on a 100 mm visual analogue scale (from worst to best).

5.3 Experiment 5 results

All data were analyzed using SPSS version 19.0 with a two-tailed *p*<.05 threshold for significance. All reaction time (RT) results are reported in milliseconds (msec). Extreme fast (<300 msec) RTs were removed from the data prior to analysis (~1%).

5.3.1 Group equivalency

Training

For training performance a 2x2 mixed-effects ANOVA was conducted using the between-participant variable of Delay type (Wake vs. Sleep) and the within-participant variable of List (A-B vs. A-C) using the dependent-variable of training blocks completed (between 4-10). This analysis yielded no main effect of Delay type, *F*<1 (*Wake* = 4.2, *SE*=.1; *Sleep* = 4.2, *SE*=.97), or List, *F*<1 (*A-B* = 4.3, *SE*=.096; *A-C* = 4.1, *SE* =.085) or interaction between these two factors, *F*<1, indicating that all participants exited training both equivalently and relatively quickly (*grand mean* = 4.2 blocks, *SE*=.07).

Accuracy levels on the final block of testing were entered into a 2x2x2 mixed-effects ANOVA featuring the between-participant variable of Delay type (Wake vs. Sleep) and within-participant variables of Congruency (Congruent vs. Incongruent) and List (A-B, vs. A-C). The products of this analysis can be found in Table 32. No main effects Delay type, *F*<1, Congruency, *F*<1, or List *F*<1 were apparent. However, a significant Delay type by List interaction was present, *F*(1,33)=5.34, *p*=.027. This interaction was driven by higher A-B list accuracy (*M*=92%, *SE*=1.6) relative to A-C list accuracy (*M*=90%, *SE*=1.6) in the Wake condition, while the opposite pattern was evident in the Sleep condition (*A-B*=91%, *SE*=1.5; *A-C*=95%, *SE*=1.5). Although this interaction is significant it should also be noted these differences in training performance are relatively small, only varying by 2-4% across lists between the two Delay types. No interaction between Delay type and Congruency, *F*<1, Congruency and List, or three-way interaction between the above factors was present, *F*(1,33)=2.1, *p*=.156.

Table 32. Experiment 5: Congruent and Incongruent accuracy levels for list learning on the final block of training.

Congruency	List type	Wake condition	Sleep condition
Congruent	A-B	93% (1.9)	91% (1.9)
	A-C	89% (1.7)	96% (1.7)
Incongruent	A-B	91% (2)	90% (1.9)
	A-C	91% (2)	94% (2)

Note: Values in parentheses denote the standard error.

Self-reported sleep measures

To assess the relative sleep patterns across the two time delays the self-reported levels of sleep both obtained normally (i.e. the average amount obtained over the two weeks prior to the experiment) and on the experimental night (i.e. prior to the testing phase) were entered into a 2x2 mixed-effects ANOVA featuring the between-participant variable of Delay type (Wake vs. Sleep), and the within-participant variable of Sleep period (Normal night vs. Experimental night) using the dependent variable of sleep amount (in hours).

A main effect of Delay type was present, *F*(1,33)=4.72, *p*=.037, indicating that, overall, the Wake condition engaged in shorter periods of sleep than participants in the Sleep condition (*Wake* = 7.6 hours, *SE*=.14; *Sleep* = 8 hours, *SE*=.13). No main effect of sleep period was apparent, *F*(1,33)=1.12, *p*=.302 (*Normal night* = 7.9, *SE*=.11; *Experimental night* = 7.8, *SE*=.12), and, critically, no interaction between Delay type and Sleep period was found, *F*<1. As such, although the sleep condition obtained more sleep in general, sleep patterns on the experimental night were not different from those obtained regularly for either condition.

To measure sleep quality subjective ratings of the night of sleep obtained using a visual analogue scale (0-10 cm, worst to best) were compared using a t-test between Delay types. A significant difference emerged, with participants in the Sleep condition reporting lower levels of sleep quality on the experimental night (*M*=5.8, *SE*=2.04) relative to the wake condition (*M*=7.64, *SE*=1.64). This difference most likely reflects a 'first night effect' due to participant's exposure to the laboratory setting and novel sleeping environment (Agnew, Webb, & Williams, 1966).

Subjective alertness levels

Alertness levels throughout the experiment were compared between the two time Delays using the self-reported measures obtained on the SSS (Hoddes et al., 1973). SSS ratings (1-7, most alert – least alert) were used as the dependent variable in a 2x2 mixed-effects ANOVA featuring the between-participant variable of Delay type (Wake vs. Sleep) and the within-participant variable of experimental Phase (Training vs. Testing). The products of this analysis can be found in Table 33.

No main effect of Delay type was evident, *F*(1,33)=1.03, *p*=.318, indicating that the overall levels of alertness across training and test were equivalent between the two time delays. A marginal main effect of Phase emerged, *F*(1,33)=2.95, *p*=.095, suggesting that, on average, participants showed a tendency to be more alert at testing than at training. A strong significant interaction between Delay type and Phase emerged, *F*(1,33)=13.97, *p*=.001. Visual inspection of the data in Table 33 suggests that, relative to the Wake condition, participants in the Sleep condition were less alert at training and more alert at testing. This difference highlights a potential confounding influence of alertness across the time delay manipulation.

Table 33. Experiment 5: Stanford Sleepiness Scale (SSS) measures.

Phase	Wake	Sleep
Training	2.1 (.17)	3.1 (.16)
Testing	2.5 (.26)	1.9 (.26)

Note: Values in parentheses indicate the standard error.

Objective alertness levels

A more objective assessment of alertness was obtained through measures of performance on the RT control task, which was implemented prior to training and testing phases (as in Figure 30). Accuracy and accurate RT measures from this task were entered into a 2x2 mixed-effects ANOVA using the between-participant variable of Delay type (Wake vs. Sleep) and the within-participant variable of experimental Phase (Training vs. Testing). The products of these analyses can be found in Table 34.

For accuracy levels no main effect of Delay type, $F(1,33)=2.67$, $p=.112$, or main effect of Phase, $F(1,33)=2.19$, $p=.148$, were apparent. Moreover, these factors did not significantly interact, $F<1$. For accurate RTs no main effect of Delay type was present, $F<1$, however a strong main effect of Phase emerged, $F(1,33)=16.54$, $p<.001$, indicative of faster RTs at test than at training. This main effect is most likely the result of practice effects as it was the second time all participants had encountered the task. Importantly, an interaction between Delay type and Phase emerged, $F(1,33)=7.57$, $p=.010$. Visual inspection of the RT data in Table 34 indicates that speed increase in RTs between training and testing was larger for participants in the sleep condition (~37 msec) relative to the wake condition (~8 msec). In conjunction with the subjective measures of alertness above, these differences are a cause for concern, and are addressed in the following section.

Table 34. Experiment 5: Accuracy (%) and Accurate RTs (msec) on the rapid response control task.

	Wake		Sleep	
Phase	Accuracy (%)	Accurate RTs (msec)	Accuracy (%)	Accurate RTs (msec)
Training	99% (.7)	334 (9)	99% (.7)	341 (10)
Testing	97% (.7)	326 (7)	99% (.8)	304 (7)

Note: Values in parentheses indicate the standard error.

5.3.2 Dealing with differences in alertness

Taken together the objective and subjective measures of alertness suggest potentially important differences between the two experimental time delays. Namely, participants in the sleep condition appear to be more alert and faster on baseline measures of RT performance at test, relative to participants who remained awake. These differences potentially reflect time of day effects, with the later time of testing in the wake condition (9-10 pm) inducing relative levels of fatigue compared to the morning testing phase in the sleep condition (9-10 am) (Borbély & Achermann, 1999). A related possibility is that participants in the sleep condition are generally more rested, and thus more alert, given the night of preceding sleep. These factors, independently or in combination, may influence performance on the primary behavioural measures featured in this investigation.

It should also be noted that the average SSS ratings did not correlate with these RT measures (unlike Experiment 5), *r*(35)=.058, *p*=.74. Nevertheless, as the RT task provides a more objective level of alertness it will be used to establish the influence of alertness on performance in analysis. As such the accurate RTs obtained from the RT task will be used as a covariate in some of the following analyses. The use of this dependent variable in an ANCOVA will allow us to see if differences on the baseline RT measure account for any behavioural differences on the primary measures of memory performance (Field, 2009, p. 397). However, Delaney and Maxwell (1981) note that the introduction of a covariate in this manner can alter the subsequent main effects due to changes in the sum-of-square errors of a given analysis (p. 107). To counteract this effect Delaney and Maxwell suggest that any covariate should be *mean centred* with the group mean of a given covariate (in this case baseline RT performance at test) deducted from the individual measures (i.e. individual RTs – mean sleep or wake condition RT). This adjustment ensures the variability of the covariate is preserved while the group means are reduced to zero, thus eliminating the impact of the covariate on the sum-of-squares error. The ANCOVAs reported in the following sections use mean centred RT values adjusted in this manner.

5.3.3 Cued recall performance across A-B and A-C lists

As with Experiments 3 and 4, cued recall in this experiment was probed by presenting participants with all 20 trained novel word items (for example ‘shusk’) and requesting that participants recall the affiliated B-list and C-list items (which for ‘shusk’ could be ‘rat’ and ‘flea’ respectively). As with the previous chapter, all possible list responses were coded in this condition (see section 4.3.2) with each % value referring to the % of total possible responses for a given response permutation. As context is of particular interest to the present analyses (i.e., % of items recalled in the correct or incorrect list) the results in this section deal with items correctly or incorrectly allocated to the B or C-lists.

To reiterate, correct list recall would mean participants correctly allocated a given real word to its original trained list (as with the example above), whereas incorrect list recall would mean that participants have recalled a given novel word in the incorrect list. Again, using the example novel word ‘shusk’, this might involve incorrectly allocating ‘rat’ to the C-list and ‘flea’ to the B-list. As with the equivalent measures in Chapter 4, correct-list recall is most directly relatable to cued recall measures obtained in previous investigations of this type (cf. Abel & Bäuml, 2013; Drosopoulos, Schulze, Fischer, & Born, 2007), whereas the incorrect-list recall provides a measure of conflation between the B and C lists.

Correct list recall

Responses coded as being in the correct list (regardless of novel word affiliation) were entered into a 2x2x2 mixed-effects ANCOVA featuring the between-participants variable of Delay type (Immediate vs. Delayed) and the within-participants variables of List (B list vs. C list) and Congruency (Congruent vs. Incongruent). As outlined in section 3.3.2 the covariate in this analysis relates to baseline RT performance at test, and the products of this analysis can be found in Table 35.

Here, a main effect of Delay type was evident, $F(1,32)=5.44$, $p=.026$, $\eta_p^2=.626$, which did not share a relationship with baseline levels of RT performance, $F<1$, indicating lower correct list recall in the Wake condition ($M=33\%$, $SE=3.6$) relative to participants in the Sleep condition ($M=44\%$, $SE=3.5$). No main effects of List, $F<1$, or Congruency, $F<1$, were forthcoming. No additional interactions were found between Delay type and List, $F(1,32)=1.88$, $p=.180$, Delay type and Congruency, $F<1$, Congruency and List, $F<1$, or three-way interaction between these factors, $F<1$. As with the equivalent measures obtained in Chapter 4, the classic RI effect is therefore not apparent in this analysis.

Table 35. Experiment 5: Recall rates for correct and incorrect list responses and untrained intrusions rates.

		Wake		Sleep	
Analysis type	List	Congruent	Incongruent	Congruent	Incongruent
Correct list responses	A-B	36% (5.7)	36% (5)	44% (5.5)	43% (5)
	A-C	32% (5)	25% (5)	44% (4.9)	47% (4.8)
Incorrect list responses	A-B	27% (3.2)	25% (3.8)	9% (3.1)	12% (3.7)
	A-C	19% (3.8)	15% (3.1)	14% (3.7)	9 % (3.5)
Untrained intrusions	A-B	2% (1.3)	7% (2.1)	6% (1.4)	3% (2.1)
	A-C	2% (1.5)	6% (1.9)	3% (1.5)	4% (1.9)

Note: List responses pertain to cued recall in the correct or incorrect list (B or C) regardless of novel word affiliation (see text). Values in parentheses represent the standard error.

Incorrect list recall

Responses coded as being in the incorrect list were entered into the same 2x2x2 mixed-effects ANCOVA as above. This analysis yielded a main effect of Delay type, $F(1,32)=8.99$, $p=.005$, $\eta_p^2=.212$, a difference that was not influenced by the baseline RT covariate, $F(1,32)=1.37$, $p=.250$. This main effect reflected higher amounts of incorrect List recall in the Wake condition ($M=21\%$, $SE=2.5$) relative to the Sleep condition ($M=11\%$, $SE=2.4$). No main effect of Congruency was present, $F<1$, however a marginal main effect of List was evident, $F(1,32)=3.23$, $p=.082$, indicative of a greater tendency for incorrect list recall in the A-B list ($M=18\%$, $SE=2.06$) relative to the A-C list ($M=14\%$, $SE=2.09$). This trend was not influenced by the baseline RT covariate, $F<1$, and was qualified by an interaction between Delay type and List, *F*(1,32)=4.71, *p*=.038. This interaction, visible in Figure 31, was driven by higher levels of incorrect list recall for A-B list items relative to A-C list items after 12 hours of wakefulness. The nature of this interaction suggests that RI effects are indeed present on this measure, but only for participants in the Wake condition who appear to be making more interlist errors for B-list, but not C-list, items. In the Sleep condition levels of incorrect list recall were not only lower, but also more closely matched between the A-B and A-C lists. No additional interactions were found between Delay type and Congruency, *F*<1, Congruency and List, *F*(1,32)=1.85, *p*=.183, or three-way interaction between the above factors, *F*(1,32)=1.05, *p*=.314.

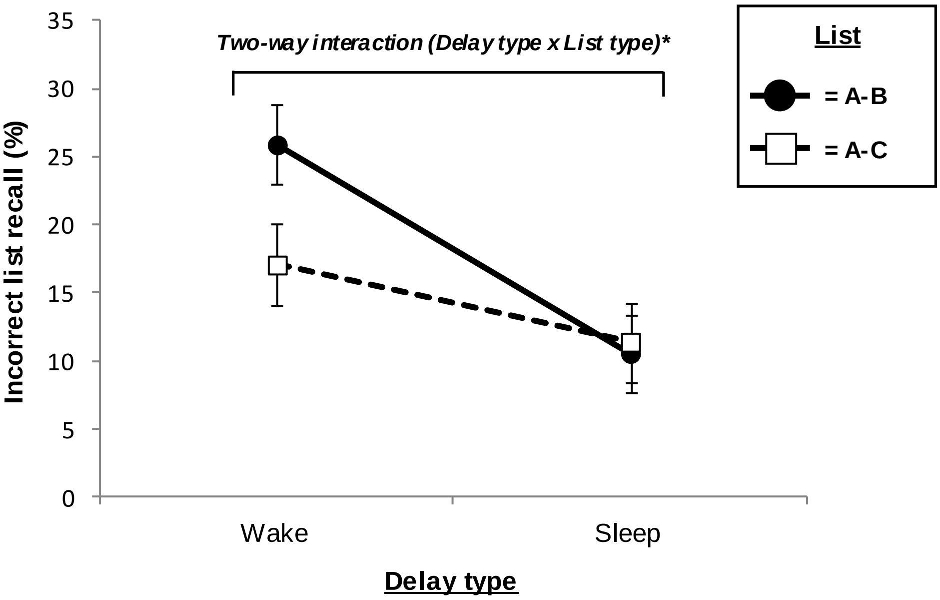


Figure 31. Experiment 5: An interaction between the levels of incorrect list recall between Delay type and List. Significance value (*) indicates *p*<.05. Error bars indicate the Standard Error.

Intrusions

Lastly, the level of untrained intrusions (e.g. ‘badger’) were entered into the same 2x2x2 mixed-effects ANCOVA as above, and the products of this analysis can also be found in Table 35. No main effects of Delay type, *F*<1, List, *F*<1, or Congruency, *F*(1,32)=2.13, *p*=.154, were present. Furthermore, no interactions were found between Delay type and List, *F*<1, or Congruency and List, *F*<1, however a marginal Delay type by Congruency interaction was present, *F*(1,32)=3.68, *p*=.064. This marginal interaction was driven by a tendency for the Wake condition to display a lower number of intrusions in the A-B list (*M*=2%, *SE*=1.11) relative to the A-C list (*M*=6%, *SE*=1.8), while error rates across lists remained closely matched in the Sleep condition (*A-B*=4%, *SE*=1; *A-C*=4%, *SE*=1.7). No three-way interaction was present. Despite the lack of robust differences across the two time delays it should be noted that the baseline RT covariate displayed a marginally positive trend between slower RTs (i.e. lower potential alertness) and the number of intrusions, *F*(1,32)=4.07, *p*=.052, *r*=.336. As such, participants with slower baseline RTs showed a tendency to recall more untrained intrusions on this measure.

Taken together these findings suggest that ongoing waking interference will result in lower levels of context-specific recall when compared to the sleep condition. These differences between the time delays do not extend to absolute differences in the amount of items recalled on this measure. A t-test between the Wake and Sleep conditions combining the total % of items recalled (regardless of List or Congruency for all coded responses) indicate almost identical levels of overall recall, *t*<1 (*Wake*=58%, *SE*=4.1; *Sleep*=59%; *SE*=5.1). This latter observation suggests that the present findings reflect a qualitative difference context-specific recall over sleep and wakefulness, as opposed to quantitative differences in the capacity for absolute recall on this measure.

Lastly, the higher % of incorrect list recall specific to the A-B list measure after 12 hours of wakefulness indicates that these items may be particularly vulnerable to waking interference. As seen in Figure 31 the levels of incorrect list recall after a period of sleep were not only lower, but also rendered more comparable between the A-B and A-C lists. This is reminiscent of the reduction in RI effects across time typically found on this measure (Underwood, 1948a, 1948b), including the 2AFC list judgements in Experiments 3 and 4. This latter finding adds additional support to a passive-protective account of sleep; in that RI effects appear to be minimised rather than recovered on this measure.

5.3.4 2AFC size judgements and retroactive interference

The second measure of interest probed the memory of the relative size relationships established across the A-B and A-C lists during training. Memory of training was assessed via a 2x2x2 mixed-effects ANCOVA featuring the between-participants variable of Delay type (Wake vs. Sleep) and the within-participants variables of Congruency (Congruent vs. Incongruent) and List (A-B vs. A-C) and the covariate of baseline RT performance. The dependent variables featured in these analyses relate to accuracy (%) and RTs (msec). RTs, as opposed to accurate RTs, were used in the latter analysis due to the chance levels of performance on some of these measures (see below).

Accuracy levels

The products of this analysis, which are presented in Figure 32, revealed two principle main effects. Firstly, a main effect of Delay type emerged with regards to overall levels of accuracy, *F*(1,32)=6.89, *p*=.013; a difference driven by higher accuracy levels in the Sleep condition (*M*=63%, *SE*=1.9) relative to the Wake condition (*M*=56%, *SE*=1.98). Secondly, a strong main effect of Congruency was present, *F*(1,32)=75.86, *p*<.001. This difference was driven by higher levels of accuracy in the Congruent condition (*M*=68%, *SE*=1.55) relative to the Incongruent condition (*M*=51%, *SE*=1.93). No main effect of List was present, *F*(1,32)=2.23, however there was a marginally significant interaction between Delay type and List, *F*(1,32)=3.83, *p*=.059, η_p^2=.107 (discussed below). Neither of the above main effects shared a relationship with the baseline RT covariate, *Fs*<1.

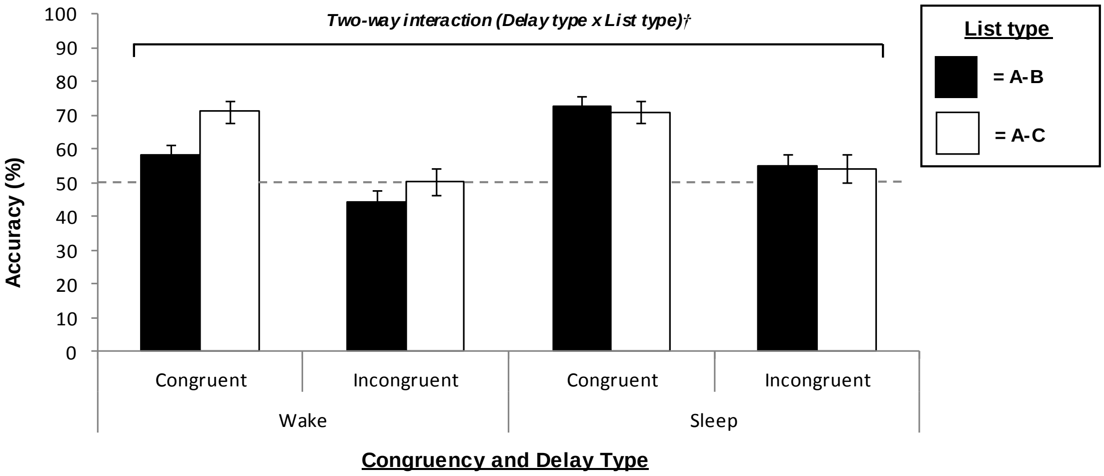


Figure 32. Experiment 5: 2AFC Accuracy levels at test across Congruency, List type and Delay type. Significance value (*†*) indicates a marginal (*p*=.059) interaction (see text). The grey dashed line in this figure indicates chance levels of responding (50%). Error bars represent the standard error

The above marginal interaction appears to be driven by a reduction in interlist RI across both congruency conditions after a period of sleep. Due to the large main effect of Congruency, RI recovery was explored further using paired sample t-tests conducted independently within the Congruent and Incongruent conditions. In the Wake condition, superior A-C relative to A-B list performance was revealed for Congruent associations, *t*(16)=-2.55, *p*=.021, while a non-significant trend indicated a tendency for RI in the Incongruent condition, *t*(16)=-1.84, *p*=.085. Conversely, these interlist differences in accuracy performance were not present in the Sleep condition for either congruency condition, *ts*<1. This supplementary analysis confirms that RI effects are eliminated after a period of sleep but not wakefulness.

Importantly, the nature of the above interaction is different from the equivalent measures presented in Chapter 4. While previously RI appeared to be reduced through a reduction in A-C (relative to A-B) accuracy over time, here RI recovery is marked by resurgence in A-B list accuracy only. This was confirmed using t-tests across the time manipulation with higher levels of A-B list accuracy in the Sleep condition confirmed for both Congruent, *t*(33)=3.46, *p*=.002, and Incongruent A-B list responses, *t*(33)=2.43, *p*=.032, while A-C accuracy levels remained equivalent across the two time delays for both forms of congruency, *ts*<1. Taken together these results are in line with an active account of sleep-associated memory consolidation given that it is the A-B list associations that appear to recover after a night of sleep (Drosopoulos, Schulze, et al., 2007; Ekstrand, 1967).

Lastly, it should be noted that the majority of responses in the Incongruent condition were at chance (~50%) levels for both Delay types, as confirmed through a series of one-sample t-tests. In the Sleep condition both A-B, *t*(17)=1.27, *p*=.220, and A-C responses, *t*(17)=1, *p*=.330, were at chance. Similarly, in the Wake condition A-C list accuracy levels were at chance, *t*<1, while A-B list accuracy levels were actually at below chance levels, *t*(16)-2.3, *p*=.035. Together these observations show that the above main effect noted for Congruency, as with Experiments 3 and 4, is manifested through a profound reduction in the ability to accurately remember specific Incongruent size relationships at both experimental time points.

RT performance

RT analysis, the products of which can be found in Table 36, did not reveal any significant main effects for Delay type, *F*(1,32)=1.14, Congruency, *F*<1, or List, *F*<1, nor did the baseline RT covariate relate to any of these comparisons, *All F*s<1. No interactions were found between Delay type and List, *F*<1, Delay type and Congruency, *F*<1 Congruency and List, *F*<1, or three-way interaction between these factors. As such there was little to distinguish any of the independent variables in this analysis.

Table 36. Experiment 5: 2AFC reaction time measures.

	Wake		Sleep	
	Congruent	Incongruent	Congruent	Incongruent
A-B List	1130 (60)	1142 (54)	1224 (58)	1231 (53)
A-C List	1150 (58)	1158 (62)	1213 (57)	1242 (60)

Note: Values in parentheses indicate the standard error.

5.3.5 2AFC generalisations and novel word size focus

The third and final behavioural measure relates to the generalisation and size focus of novel word items in relation to new but familiar real word items. The factor of 'Size focus' was formed by performing a median split along the range of rated size distances between their affiliated real word items (ranging from .6 to 6.15). Novel words trained between real words with a more concentrated size range (rated differences ranging from .6 to 2.39) were designated as 'Narrow', while novel words trained between real words with a broader size range (ranging from 2.65 to 6.15) were designated as 'Wide'. As with the previous two experiments, congruent and incongruent responses were compared in the same analysis, with size judgements falling within the size limits established in training marked with a positive % value, and those falling outside trained limits allocated a negative % value, as part of a collective variable labelled 'generalisation %'.

Generalisation % Analysis

A 2x2x2 mixed-effects ANCOVA was run using the between-participant variable of Time delay (Wake vs. Sleep) and the within-participant variables of Congruency (Congruent vs. Incongruent) and Size Focus (Narrow vs. Wide) with the covariate of baseline RT performance at test (msec). The products of this analysis can be found in Table 37.

Table 37. Experiment 5: Generalisation performance in the wake and sleep conditions.

		Wake		Sleep	
Congruency	Size focus	Generalisation (%)	RTs (msec)	Generalisation (%)	RTs (msec)
Congruent	Narrow	69 (2.4)	1055 (57)	69 (2.3)	1115 (56)
	Wide	62 (1.9)	1033 (57)	66 (1.9)	1130 (56)
Incongruent	Narrow	71 (1.9)	1050 (58)	67 (1.9)	1113 (57)
	Wide	63 (2.6)	1054 (54)	68 (2.6)	1098 (52)

Note: Narrow and Wide focus items are based on a median split of the rated size distance between two real words (see text). Values in parentheses represent the standard error.

No main effects of Delay type, *F*<1, or Congruency, *F*<1 were present, however a main effect of Size focus did emerge, *F*(1,32)=5.69, *p*=.023, indicative of higher levels of generalisation % for Narrow (*M*=69%, *SE*=1.38) as opposed to Wide focus size judgements (*M*=65%, *SE*=1.33). This main effect, which did not share a relationship with the covariate, *F*<1, suggests that Narrow size novel words are provided a more discerning basis for generalisations at test. A marginally significant interaction between Delay type and Size focus was present, *F*(1,32)=3.64, *p*=.065. This interaction shows that participants in the Wake condition displayed a tendency for higher levels of generalisation for Narrow (*M*=70%, *SE*=1.97) as opposed to Wide size focus novel words (*M*=63%, *SE*=1.9). However, in the Sleep condition generalisation levels appear more equivalent for Narrow (*M*=68%, *SE*=1.9) and Wide (67%, *SE*=1.9) focus generalisations. No further interactions were present between Delay type and Congruency, *F*<1, Congruency and Size focus, *F*<1, or three-way interaction between the above factors.

To supplement the above analysis the factor of Size focus was broken down into its component units of Size focus (ranging from .06 to 6.15) which were then independently correlated with accuracy performance for the Wake and Sleep conditions using Pearson correlation coefficients. Congruent generalisations (presented in the top half of Figure 33) revealed a significant negative relationship between these variables in the Wake condition only, *r*(170)=-.229, *p*=.003, indicating that novel word items with an increasingly narrow focus display higher levels of generalisation % after a period of wakefulness. In the Sleep condition there was a non-significant trend in the same direction for Congruent generalisations, *r*(180)=-.110, *p*=.160. However, when the magnitude of these two coefficients was compared using a Fisher's r-to-z transformation it was found that they were not significantly different between the two time delays, Z=-1.12, *p*=.25. Moreover, it should be noted that on closer inspection of the congruent correlations (presented in Figure 33), the low generalisation scores associated with the largest Size focus unit (6.15) may be outliers, driving the above negative correlation in the Wake condition. Indeed, removal of this size focus unit from analysis rendered the above correlation in the Wake condition non-significant, *r*(160)=-.090, indicating that this relationship is not particularly robust.

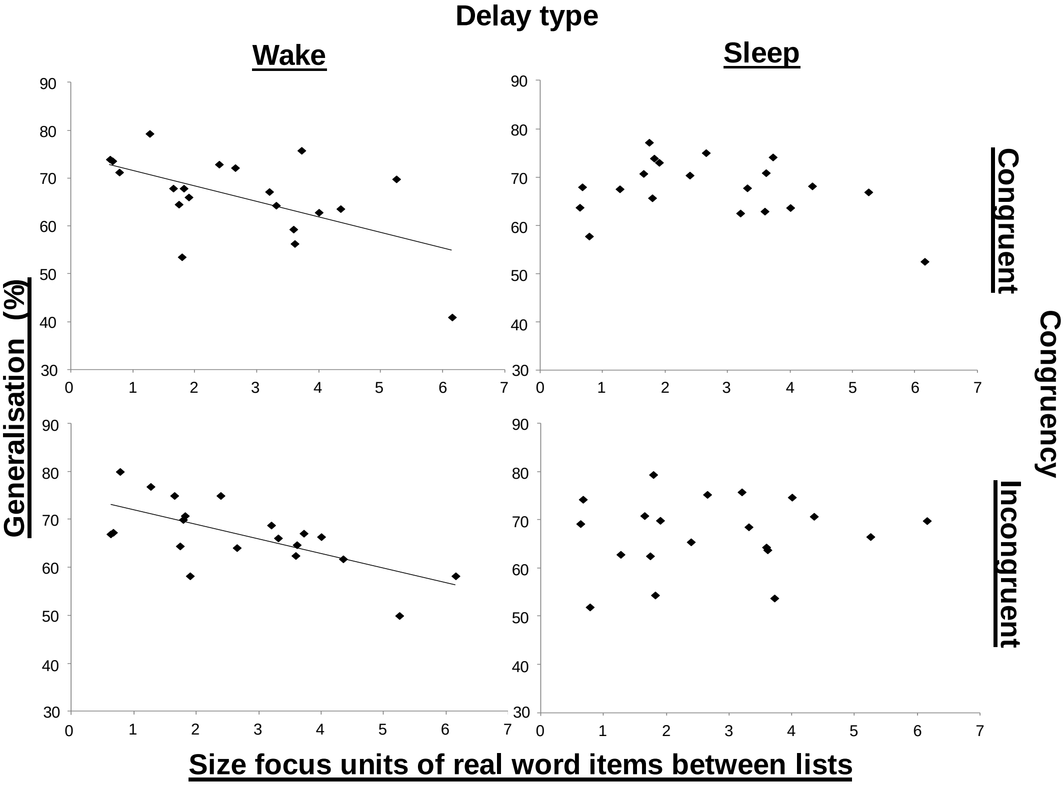


Figure 33. Experiment 5: Generalisation levels (%) correlated with the differences in size focus between real word items for both congruent and incongruent responses across the 24 hour time delay. Note that generalisation scores in this figure are averaged across each size focus unit.

Incongruent measures (presented in the bottom half of Figure 33) again revealed a negative relationship between Size focus and generalisation scores in the Wake condition, *r*(170)=-.277, *p*<.001, while no such relationship was present in the Sleep condition, *r*(180)=.037, *p*=.621. A Fisher's r-to-z, indicated that the magnitude of this relationship was significantly greater in the Wake condition relative to the Sleep condition, Z=2.29, *p*=.022. It should also be noted that, unlike the Congruent responses, the significant correlation in the Wake condition for Incongruent generalisations was not contingent on generalisation scores for the most distal size focus unit (6.15). This point is exemplified in an equivalent negative correlation in the Wake condition when analysis was repeated once this distant size focus unit was removed, *r*(163)=-270, *p*<.001. As the impact of size focus was only expressed after a period of wakefulness, these data are more in line with the passive protective account of sleep outlined in the predictions for this measure.

Generalisation RT analysis

RTs for the generalisation measures were entered into the same 2x2x2 mixed-effects ANCOVA used to analyse the generalisation % scores in the previous section, and the products of this analysis can also be found in Table 37. No main effects of Delay type, *F*<1, Congruency, *F*<1, or Size focus, *F*<1, were present, nor did the baseline RT covariate interact with any of these main effects, All *F*s<1. No Interactions were present between Delay type and Size focus, *F*<1, Delay type and Congruency, *F*<1, Congruency and Size focus, F<1, or three-way interaction between any of these factors, *F*<1. As such, RT measures were not differentiated by any of the variables entered into this analysis.

5.3.6 Sleep stage correlations

The data from the 15 intact sleep records, presented in Table 38, show sleep parameters within the range expected from a sample of normal, healthy young adults (Spriggs, 2008).

Table 38. Experiment 5: Sleep parameters.

Sleep parameter	Mean time (min) ± SEM	% of total sleep time ± SEM
Total sleep time	487 ± 7.2	
Sleep latency	35 ± 5.6	
Stage 1	69 ± 1.2	14 ± 0.3
Stage 2	242 ± 9.4	50 ± 1.5
Slow Wave Sleep	101 ± 9.4	21 ± 1.4
REM sleep	116 ± 5.2	24 ± 1

Note: Sleep parameters in parentheses represent the relative sleep stages as defined by Rechtschaffen and Kales (1968) excluding Slow Wave Sleep which is stages 3 and 4 collapsed. REM = rapid eye movement, SEM = standard error of the mean.

The original intention of the present analysis was to simplify the behavioural measures reported in the previous sections by reducing them down into aggregated factors representing the principle effects of interest. However the 2AFC list learning measures displayed such absolute recovery from RI that this composite measure produced little variation. Indeed, the majority of values on this aggregated factor were near, or close to zero. This led to a non-normal distribution of scores as indicated by a K-S test of normality for congruent responses, *D*(15)=.323, *p*<.001. To circumvent this issue a composite factor was not created (i.e. A-B list – A-C list) and instead the original measures were entered independently for both list performance and congruency. As such, a total of 8 normally distributed 2AFC list factors were entered into correlational analysis (including both accuracy and RT performance); twice the amount featured in the equivalent analyses conducted in Experiments 3 and 4.

A second issue relates to the cued recall measures, specifically the measures of incorrect list recall. As can be seen in Table 35, % scores are not only low for this measure but also do not produce sufficient variability for correlational analysis. In this case it was deemed necessary to exclude the incorrect list responses from analyses. Correct list responses, however, were viable and thus aggregated down into a single factor representing RI, as with Experiments 3 and 4 (A-B list accuracy – A-C list accuracy). This was calculated for congruent and incongruent responses to produce a total of two factors for cued recall.

Lastly, there was enough variation in the data to produce a simplified size factor derived from the size focus variables (Narrow or Wide). Accordingly, a size focus factor was created by deducting Narrow generalisation performance from Wide generalisation performance to produce a single composite value representing the extent of this difference. This factor was calculated independently for congruent and incongruent generalisations using both % performance and RT measures, producing a final four factors for correlational analysis.

The above 14 factors were then correlated with TST (mins), REM sleep %, Stage 2 sleep % and SWS %. No correlations were forthcoming between sleep physiology and any of the above behavioural measures. As such corresponding table of results has been placed in Appendix 14. The analysis was repeated using the time (in minutes) spent in REM sleep, Stage 2 sleep and SWS (Appendix 15). The only correlation to emerge on this secondary analysis was a negative relationship between the time spent in REM sleep and Incongruent A-C list accuracy, $r(15)=-.572$, $p=.023$. However this correlation did not survive correction for multiple comparisons, and thus likely represents a spurious artefact. Taken together the above analysis indicates no discernible relationship between sleep stage activity and memory performance.

5.3.7 Sleep spindle activity

Table 8 displays the spindle count and density measures for the three different frequencies extrapolated from the sleep data (presented across stage 2 and SWS). Unlike Experiment 2, the spindle values from Experiment 5 are generally higher and more comparable to those found in investigations using the same spindle detection algorithm (c.f. Tamminen, 2010).

Table 39. Experiment 5: Spindle count (total n) and density (spindles per minute).

		Spindle Count (total n)		Density (per minute)	
Spindle frequency	Sleep stage	central electrodes	frontal electrodes	central electrodes	frontal electrodes
All (11-15 Hz)	Stage 2	156 (20)	269 (27)	0.65 (.06)	1.14 (.08)
	SWS	31 (4)	51 (7)	0.32 (.04)	0.53 (.08)
Fast (13-15 Hz)	Stage 2	183 (22)	255 (28)	0.77 (.07)	1.09 (.09)
	SWS	48 (6)	65 (10)	0.49 (.06)	0.66 (.09)
Slow (11-13 Hz)	Stage 2	134 (16)	257 (25)	0.57 (.06)	1.10 (.08)
	SWS	19 (3)	38 (6)	0.20 (.04)	0.42 (.08)

Note: Standard error in parentheses. SWS = slow wave sleep. Central electrodes refer to aggregated values across electrode sites C3 and C4. Frontal electrodes refer to aggregated values across electrode sites F3 and F4.

Spindle density correlated with behavioural factors

The final stage of analysis in Experiment 5 was to correlate the 14 behavioural factors featured in the previous section with spindle density. Given the large number of potential comparisons that can be derived from the rich dataset obtained in the present experiment the broadest possible assessment of spindle activity was conducted in relation to the above behavioural factors. Following on from similar analyses conducted in previous studies (Schmidt et al., 2006; Tamminen et al., 2010) spindle density was aggregated across the entire NREM period (Stage 2 + SWS) independently for fast (13-15 Hz), slow (11-13 Hz) and all spindle activity (11-15 Hz).

The products of this analysis, which can be found in Appendix 16, revealed no significant relationships. Additional unreported analyses were also conducted factoring in electrode topography (frontal or central) and sleep stage (stage 2 or SWS). Again, no relationships of note were found (Appendices 17-19), indicating that sleep spindle activity did not share a meaningful relationship with any of the present behavioural factors.

5.4 Summary and Discussion

The aim of Experiment 5 was to elaborate on the observations made in Experiments 3 and 4 which, using a modified A-B, A-C list learning paradigm, revealed several interesting time-dependent changes to memory. However, a noted issue with the exclusive use of 20 minute and 24 hour time points in these previous investigations was that an active account of consolidation could not be distinguished from alternative accounts implicating memory decay over wakefulness. To investigate this matter, Experiment 5 took advantage of a 12:12 sleep/wake design to probe the relative influences of sleep and wakefulness on memory within a 24 hour period. To this end, Experiment 5 extended the measures obtained from Experiment 3, which featured the same training criterion and stimuli as the present investigation. High-quality PSG recordings were also implemented to explore the potential physiological correlates of memory consolidation. The following summary of the key findings from this experiment are outlined in terms of the active vs. passive-protective hypotheses regarding sleep framing this investigation.

An active account of consolidation was largely corroborated in memory performance relating to the size relationships established across the A-B and A-C word lists. Specifically, there was evidence for superior A-B list memory after a period of sleep relative to wakefulness. This pattern of performance runs counter to the prediction that waking interference alone will eliminate the influences of RI through the more prominent forgetting of A-C list size relationships (Slamecka & Ceraso, 1960). Indeed, counter to this position, not only were RI effects preserved exclusively in the wake condition but accuracy levels for the A-C list size relationships remained equivalent across both time delays. These results instead verify the hypothesis that the sleep state can facilitate the recovery of size relationships that have been detrimentally influenced by RI (Drosopoulos, Schulze, et al., 2007). However, despite this interesting finding, there was no evidence for a relationship between these behavioural effects and sleep physiology itself. This null result suggests that sleep may aid mnemonic recovery through other, as yet unspecified, processes.

A passive-protective account of the sleep state applies to the findings obtained from the other two behavioural measures in this investigation. Firstly, recall after 12 hours of wakefulness resulted in a greater number of interlist intrusions, and a concurrent drop in levels of list-specific recall, relative to the sleep condition. These observations suggest that sleep may protect context-specific memory from the deterioration otherwise seen after an equivalent period of wakefulness. Secondly, real word size relationships established across the B-A-C word triplets did not influence the formation of novel word generalisations after a period of sleep. Instead, generalisation performance relating to novel words with a less established locus of size deteriorated after 12 hours of wakefulness; an effect that did not extend to items with a more clearly defined size range during this time. Again, this result suggests that sleep can help protect memories that are prone to waking decay. Lastly, a passive-protective role for sleep in relation to both of these findings was indirectly supported through the null results regarding sleep physiology. This latter observation further suggests that sleep did not actively influence memory performance on these measures.

5.4.1 Spontaneous recovery and memory consolidation

In the predictions made for this experiment two competing hypotheses were introduced to explain the reduction of RI effects for the size relationships established across the A-B and A-C word lists. The behavioural predictions made in these two accounts usefully distinguished between active and passive-protective roles of sleep on memory consolidation. To reiterate, a passive-protective account would predict the relative decay of A-C list memories over wakefulness, as it is these size relationships that are deemed to be the most susceptible to waking interference (Slamecka & Ceraso, 1960). Conversely, an active account of sleep and memory consolidation would predict the spontaneous recovery of A-B list relationships in this condition through the reactivation and reprocessing of weakened memories during SWS (Drosopoulos, Schulze, et al., 2007; Melton & Irwin, 1940).

The evidence from this particular measure supports the latter hypothesis, as RI effects remained preserved between the A-B and A-C lists after 12 hours of wakefulness, with no evidence for diminished A-C list accuracy at this time point. However, after a period of sleep RI effects were not only absent, but A-B list accuracy was ~10% higher relative to the wake condition, indicating that memory for these size relationships had recovered during this time (Melton & Irwin, 1940). This consolidation-associated effect has, to the authors knowledge, only been observed for B-A-C word triplets devoid of any overlapping semantic or contextual features (e.g. '*book-egg-dog*') (Drosopoulos, Schulze, et al., 2007; Ekstrand, 1967). The present investigation demonstrates that the equivalent phenomenon can occur for the semantic associations established between trained word lists. Indeed, this novel finding is further corroborated by the absence of such effects in the cued recall measures, which do not directly relate to the memory of these interlist relationships.

Interestingly, however, RI recovery appears to be present for both forms of congruency despite the large impact of this manipulation on memory. Unlike the retention of congruent size relationships, incongruent list memory performance was predominantly expressed at chance levels over both experimental time points. Moreover, the retention of incongruent size relationships did not relate to SWS activity, an observation which would have otherwise supported an active role for sleep in the ongoing suppression of potentially superfluous memories (Payne et al., 2009; Tononi & Cirelli, 2003, 2006). Taken together these results indicate that sleep was not involved in the retention of incongruent information, resulting in a parallel drop in memory performance across both the sleep and wake conditions. Importantly, despite these low levels of accuracy there was still evidence for spontaneous recovery in this condition as incongruent A-B size relationships also showed a ~10% improvement after a period of sleep relative to wakefulness. The observation that incongruent information can recover after a period of sleep, albeit to chance levels of performance, has ramifications regarding the role that sleep may have in relation to this effect.

Specifically, the literature to date has attributed spontaneous recovery during sleep to the systems-level interaction between memory systems during this state (Drosopoulos, Schulze, et al., 2007). Here, the reprocessing of hippocampally held memories during SWS is thought to be undifferentiated between stronger and weaker memory traces, resulting in the relatively enhanced signal transfer of memories retained at a sub-optimal thresholds to the neocortex (Maquet, 2000). Importantly, this account implicates SWS in both the general recuperative benefits of sleep on memory (see Diekelmann et al., 2010; Schmidt et al., 2006) as well as the recovery of memories weakened through the more specific influences of RI. Although RI recovery was indeed observed after a period of sleep, in the present investigation this account is not entirely parsimonious when applied to the present dataset. Firstly, incongruent memories appear to have been minimally influenced by systems-level consolidation, yet this account proposes that spontaneous recovery is the product of these very same processes. Secondly, no relationships were expressed between mnemonic recovery and aspects of sleep physiology thought to reflect systems-level consolidation.

Taken together these data suggest that RI recovery may instead be attributed to a set of processes during sleep distinct from the general systems-level account outlined by Drosopoulos et al. (2007). With regards to the contributions of sleep itself it should also be noted that despite obtaining PSG sleep recordings in their investigation, Drosopoulos et al. (2007) failed to report any functional relationship between sleep physiology and their findings. As such, despite the repeated observation that sleep can indeed facilitate RI recovery (Abel & Bäuml, 2013; Drosopoulos, Windau, et al., 2007; Ekstrand, 1967), the specific contributions of the sleep state in relation to this effect remain elusive. Importantly this contention does not rule out the possibility that other aspects of the sleep state may be contributing to this behavioural outcome.

Two aspects of sleep physiology not directly measured in this investigation deserve further exploration in this regard, particularly given the potential low levels of power in the PSG analyses for this investigation (see Appendix 21). Firstly, Norman, Newman, and Perotte (2005) propose that theta-wave oscillations occupying the 4-8 Hz frequency range during sleep may actively strengthen weakened A-B list associations while simultaneously ‘punishing’ competing A-C list memories. Although theta- wave activity is expressed predominantly in REM sleep (see Diekelmann & Born, 2010), which did not correlate with mnemonic recovery in the present investigation, this does not preclude its potential contribution to declarative memory consolidation across the entire sleep period (Schabus, Hödlmoser, Pecherstorfer, & Klösch, 2005). Future investigations of this type would benefit from a more comprehensive exploration of this frequency band and its relationship with RI recovery. A second candidate mechanism relates to hippocampal sharp wave-ripples, bursts of high-frequency activity in the 100-200 Hz range, which have been closely associated with the spontaneous reactivation of experience-specific hippocampal memories (Shen & McNaughton, 1996; Wilson & McNaughton, 1994). Given the positive association between hippocampal activity and the resistance of A-B associations to forgetting (Kuhl et al., 2010) sharp-wave ripples localised in this region during sleep may also contribute to the subsequent recovery of these associations.

5.4.2 The role of forgetting and wakefulness on memory

Although the memory of interlist relationships suggests an active contribution of sleep on memory consolidation, the cued recall and generalisation measures indicate patterns of behaviour more in line with a passive-protective account of this state. Firstly, a 12 hour period of wakefulness but not sleep was associated with a clear deterioration of context-specific memory in the cued recall measure, as registered by an increased number of inter-list errors and a simultaneous drop in correct-list recall. Similarly, size judgements featuring novel word items with a broader (and thus less specific) trained size appear to be particularly vulnerable to decay over 12 hours of wakefulness. This latter finding was illustrated through the drop in generalisation % performance for these items across both forms of congruency. Lastly, a passive-protective account was more indirectly confirmed, albeit through a null set of results, by the lack of a meaningful relationship between sleep physiology and memory retention in the sleep condition.

The changes to memory observed after a period of wakefulness in the cued recall and generalisation measures are most parsimoniously explained by the relatively simple interpretation that sleep has protected these memories from the deleterious effects of waking interference. This contention is by no means a new one as highlighted by some of the first published studies in this area. This includes work by Jenkins and Dallenbach (1924) who first showed that the temporal forgetting curve for the memories of nonsense syllables across wakefulness could be impeded by periods of sleep. Memory decay in this manner may be attributable, in part, to the ongoing reliance on the hippocampus and its related structures to encode new memories across wakefulness (Dudai, 1996; Martin, 1999). Indeed the very act of being awake may negatively impact previously encoded information by placing continual demands on our resource-limited memory systems (Wixted, 2004). When this notion is applied to the present findings it would appear that sleep, in providing a respite from wakefulness, may effectively block some aspects of natural forgetting.

5.4.3 Conclusions

The aim of the present experiment was to provide further insight into the changes to memory observed over time in Experiments 3 and 4. When considering the complementary 20 minute and 24 hour time points featured in these investigations Experiment 5 provides a clearer picture of how sleep has influenced behaviour. In Chapter 4 it was suggested that the decontextualisation of memory over 24 hours in the cued recall measures may be an adaptive outcome of consolidation (Inostroza & Born, 2013; Payne et al., 2009). Moreover, it was further proposed that sleep may help actively assimilate the more useful size relationships established across B-A-C word triplets to produce the pattern of novel word generalisation found after 24 hours in Experiment 3 (Rogers & McClelland, 2004). Experiment 5 refutes these explanations by demonstrating that a period of wakefulness, but not sleep, can bring about these changes to behaviour.

It is important to note, however, that a passive-protective account of sleep should not be considered mutually exclusive from the active processes attributed to this state in Experiment 5 (Stickgold, 2005). In particular, the RI effects observed 20 minutes after training in Chapter 4 do not appear to diminish after 12 hours of wakefulness in Experiment 5. Instead sleep appears to be both sufficient and necessary for the recovery of size relationships detrimentally influenced by the acquisition of new information. Moreover, the recovery of A-B size relationships seen in Experiment 5 explains the equivalent levels of accuracy found across B-A-C word triplets after 24 hours in Experiments 3 and 4.

Lastly, as Experiment 5 has yielded additional evidence of list memory recovery over time these data also provide support for an adaptive account of sleep-associated RI recovery (Fenn et al., 2003). Here it is proposed that the acquisition of new memories throughout the waking day will often result in a degree of contextual overlap between newly formed associations, potentially weakening memories formed earlier in time. This outcome is potentially undesirable as both recent and remote memories may usefully contribute to future behaviour. Sleep, in facilitating RI recovery, may therefore help negate the cost of continued learning over periods of wakefulness.

Chapter 6: Thesis summary and conclusions

6.1 Thesis summary

The experiments reported in this thesis were designed to explore how prior knowledge may influence memory consolidation. This overarching theme was segregated into two underlying research questions. Firstly, how does information that is consistent or inconsistent with pre-existing memory differ when it comes to consolidation? Secondly, what mechanisms, particularly with regards to sleep physiology, may be associated with these processes? The twin aims of this thesis alternated between chapters which explored changes to memory that might occur over 24 hours (Chapters 2 and 4) or after a 12 hour period containing sleep or wakefulness (Chapters 3 and 5).

Moreover, the present thesis dovetailed in its approach to the principal research question. Experiments 1 and 2 featured a modified six-element transitive inference (TI) paradigm to explore how prior knowledge may influence the abstraction and assimilation of new information after consolidation opportunity. In a complementary series of investigations, Experiments 3, 4 and 5 implemented a modified A-B, A-C list learning paradigm to investigate how prior knowledge may influence the consolidation-associated recovery of memories from the detrimental effects of retroactive interference (RI). The following section will summarise the motivations and main findings of the four experimental chapters contained in this thesis.

6.1.1 Thesis motivation

In the introduction of this thesis four bodies of evidence were presented which aligned to form the basis of the experimental work conducted in Chapters 2-5. Firstly, the existence of memory consolidation was established, a cascade of molecular, cellular and system-level processes that provide useful and enduring changes to long term memory (Dudai, 2004; McClelland et al., 1995). Secondly, it was shown that prior knowledge may partly mediate the effects of consolidation. Namely, new information that displays a compatible relationship with what is already known may benefit from a ‘fast-tracked’ route of neurological and behavioural integration (Tse et al., 2007; van Kesteren et al., 2012). Thirdly, sleep was established as an ideal state for consolidation to occur as it confers a set of endogenous processes causally associated with the reactivation and reprocessing of memory (Frankland & Bontempi, 2005; Rasch & Born, 2013). Lastly, evidence was presented illustrating the relationship between consolidation and several qualitative and quantitative shifts in the expression of memory. These changes to behaviour have been revealed through a range of paradigms demonstrating the benefits of sleep in the stabilisation, enhancement, integration and recuperation of memory (Conte & Ficca, 2012;

Diekelmann & Born, 2010a; Stickgold & Walker, 2013). However, it was concluded in Chapter 1 that despite the recent and prolific developments in the existing literature a relative paucity of research has been conducted at the intersection of these complementary research agendas.

6.1.2 Chapter 2

Experiment 1 in Chapter 2 tackled the issue of how the compatibility of new information with prior knowledge might influence two forms of integration: assimilation and abstraction (Lewis & Durrant, 2011; Walker & Stickgold, 2010). Consolidation-associated abstraction has previously been demonstrated using the TI paradigm, where it has been shown that a sleep opportunity can enhance the ability to flexibility interrelate novel combinations of information extrapolated from an ordinal set of trained premises (e.g. A>B>C>D>E>F) (Ellenbogen et al., 2007; Werchan & Gómez, 2013). However, given the arbitrary nature of the trained stimuli (textured patterns) it was argued that these investigations may not fully corroborate the potential benefits of consolidation on declarative memory. Moreover, the observed patterns of integration in these studies was more indicative of procedural, rather than declarative, memory consolidation (Frank et al., 2006; von Fersen et al., 1991).

Experiment 1 aimed to provide a more definitive exploration of how consolidation might influence the integration of declarative memory using the TI paradigm. To do this training featured interleaved novel and real word items that either formed a complementary size dimension (e.g. giraffe > swock> sheep > creal > lizard > floth) or a size dimension that contained conflicting information (e.g. sheep > swock > lizard > creal > giraffe > floth). This manipulation was intended to create a set of size relationships that were either congruent or incongruent with prior knowledge. Based on recent neurological and behavioural evidence (Tse et al., 2007; van Kesteren, Fernández, et al., 2010; van Kesteren, Rijpkema, et al., 2010) it was predicted that congruent information would show more signs of mnemonic integration after 24 hours as opposed to 20 minutes. Integration was probed through three behavioural measures relating to (i) the assimilation of central premises, (ii) the abstraction of novel word TIs, and (iii) the generalisation of novel word items to new, but familiar real word items.

Two of the three measures featured in this experiment revealed evidence for enhanced memory integration over time. Firstly, relative to a 20 minute time delay, testing after 24 hours displayed a ~300 msec speed increase in the response times for central premise items (e.g. B>C>D>E). Importantly, this finding ran counter to the direction of the serial position effect, a behavioural phenomenon in the TI paradigm typically expressed through weaker retention of central premises (Greene et al., 2001; Moses et al., 2006).

Accordingly, this result suggested that memories across the entire schematic dimension had been assimilated over 24 hours. Secondly, novel word TIs (B>D inferences) were faster by a striking ~500 msec after an extended delay, indicating that the capacity to flexibly interrelate entirely novel elements of information had also improved after an extended delay.

However, despite these findings, the congruency manipulation did not influence RT integration as originally predicted. Further inspection of the data indicated that participants were consistently interpreting incongruent size relationships as congruent prior to a consolidation opportunity resulting in below chance levels of accuracy on some of the measures. This experiment was therefore argued to demonstrate two ways in which prior knowledge can influence declarative memory consolidation: firstly in facilitating the extrapolation of relevant information during or shortly after training (Stickgold & Walker, 2013), and secondly, in facilitating the swift integration of this information over 24 hours (van Kesteren et al., 2012).

6.1.3 Chapter 3

Having established that semantic information can facilitate the time-dependent integration of new information, Experiment 2 aimed to focus on the potential contributions of sleep in relation to these effects. Elements of the modified TI paradigm featured in Experiment 1 were carried forward into Experiment 2, which implemented the equivalent behavioural measures across three time points: a 20 minute time delay or a 12 hour period containing sleep or wakefulness. However, unlike Experiment 1, the incongruent condition was removed in this investigation with the aim of providing cleaner behavioural measures to correlate with sleep physiology. On the basis of the literature reviewed in Chapters 1 and 3 it was predicted that the sleep state, relative to wakefulness, would confer more benefits to mnemonic integration (Frankland & Bontempi, 2005; Rasch & Born, 2013). In particular, an association between slow wave sleep (SWS) and its affiliated spindle activity (11-15 Hz) was hypothesised to emerge in relation to the markers of behavioural integration previously seen in Experiment 1 (Durrant et al., 2011; Lau et al., 2010; Tamminen et al., 2010).

Counter to these predictions memory performance remained largely undifferentiated across the three time delays in Experiment 2. The lack of integration relative to Experiment 1 was attributed to the purely congruent training in Experiment 2, which appeared to induce faster RTs on the behavioural measures prior to a consolidation-opportunity. This acceleration in RTs potentially diminished the subsequent influence of consolidation, which has otherwise been shown to enhance more weakly retained information (Diekelmann et al., 2009; Stickgold, 2009). Despite the null set of results regarding the three behavioural measures there was provisional evidence that sleep might help maintain memories over time. Firstly, there was little deterioration in memory performance across the three time delays (Wixted, 2004). Moreover, several relationships between sleep physiology and memory retention were expressed, including weak associations between SWS spindle density and the accuracy levels of both central premises and novel word TIs. It was tentatively concluded that sleep, as well as wakefulness, may support the preservation of memories that are compatible with prior knowledge.

6.1.4 Chapter 4

In Chapter 4 it was speculated that relative differences in integration observed between Experiments 1 and 2 might be contingent on memory reactivation during consolidation. As Experiment 2 featured more congruent training than Experiment 1 any resulting memories may have failed to engage consolidation-associated processes. This contention was supported by the existing literature consistently highlighting an enhancive effect of consolidation for information retained at lower thresholds (Kuriyama et al., 2004; Schmidt et al., 2006). Of particular relevance to Experiments 3 and 4 was the A-B, A-C list learning paradigm, which provides a measure of RI effects between the memories of two sequentially learned word lists. Here, by virtue of the common ‘A’ item, primary A-B list memories will be detrimentally influenced by the acquisition of the secondary A-C list (Barnes & Underwood, 1959). In the existing literature the effects of RI have been observed to dissipate across time for word lists devoid of any implied semantic relationship (Underwood, 1948a, 1948b) a behavioural phenomenon that has been attributed to the sleep-associated consolidation and recovery of memories (Ekstrand, 1967; Ellenbogen et al., 2007).

Despite this evidence, it was also noted that the benefits of consolidation on ‘weakened’ memories must be accompanied by a degree of selectivity in order to avoid an influx of potentially maladaptive information into long term memory (Stickgold & Walker, 2013). On this basis, it was hypothesised that the applicability of newly learned information to pre-existing memory would influence its consolidation-associated recovery over 24 hours. To explore this possibility Experiments 3 and 4 featured an adapted A-B, A-C list learning paradigm with a semantic component introduced to list learning. Here, ‘A’ items were replaced with novel words, and word triplets across the two lists featured real word size relationships that were either congruent (e.g. camel > gorby > duck) or incongruent (e.g. camel < gorby < duck) with prior knowledge. Three measures were implemented to test this prediction: (i) cued recall of the real word items contained in each list, (ii) memory of the specific size relationships established across word triplets, and (iii) the generalisation of novel word items to new, but familiar real word items. Moreover, the depth of encoding was altered between Experiments 3 and 4 to a lower (≥80%) and higher (≥95%) thresholds respectively to further explore the sensitivity of consolidation for memories of differing strength (Stickgold, 2009).

As predicted the RI effects between the A-B and A-C list memories dissipated after a 24 hour delay in both Experiments 3 and 4. This time-dependent reduction in RI was observed exclusively for the memories of size relationships across word triplets and not the items themselves (i.e. cued recall). Moreover, the memory of incongruent list relationships had largely dropped to chance levels after 24 hours, suggesting that this information had been rapidly forgotten. Although the effects of RI and congruency did not extend to the cued recall measures, there was instead evidence for reduced context-specific memory over time, as participants in both experiments recalled fewer items in the correct list after an extended delay. Lastly, size generalisations were the only behaviour prominently influenced by the differences in encoding depth between Experiments 3 and 4. At a lower training threshold in Experiment 3 (≥80%) novel word generalisations reflected the usefulness of B-A-C list size training after 24 hours, yet at a higher threshold in Experiment 4 (≥95%) a similar effect emerged after just 20 minutes.

The changes to memory seen over 24 hours in Experiments 3 and 4 were shown to be compatible with the existing literature regarding memory consolidation. Specifically, sleep has been associated with the decontextualisation of episodic memories (Inostroza & Born, 2013), the spontaneous recovery of memories from RI (Drosopoulos, Schulze, et al., 2007) and the enhanced capacity for generalisation (Payne et al., 2009). Moreover, the lack of generalisation effects in Experiment 4 also aligned with the contention that consolidation is less beneficial for more strongly encoded information (Schmidt et al., 2006; Stickgold, 2009). However, an important factor limiting these interpretations was that forgetting across periods of wakefulness may have also contributed to these changes in memory performance (Slamecka & Ceraso, 1960; Wixted, 2004).

6.1.5 Chapter 5

The data obtained in Chapter 4 motivated Experiment 5, which explored two additional time points using the modified A-B, A-C list learning paradigm: a 12 hour period containing either sleep or wakefulness. These time points extended the procedures of Experiment 3 in order to distinguish the potential passive-protective and active contributions of the sleep state on memory. It was argued that if the behavioural changes seen after 24 hours in Experiment 3 emerged exclusively after a 12 hour period of wakefulness then this would support the contention that sleep had passively protected these memories from waking interference (Slamecka, 1966; Wixted, 2004). On the other hand, if the equivalent effects emerged after a period of sleep and not wakefulness, this would support an active account of consolidation during sleep (Drosopoulos, Schulze, et al., 2007; Payne et al., 2009). Measures of sleep physiology were also included in Experiment 5 under the proviso that a meaningful relationship between memory retention and sleep architecture, in particular fast (13-16 Hz) sleep spindles, would further verify an active account of this state (Schmidt et al., 2006; Wilhelm et al., 2011).

Experiment 5 found evidence supporting both the passive protective and active accounts of sleep. With regards to cued recall, context-specific list memory was found to deteriorate across periods of wakefulness but not sleep, thus corroborating the passive-protective account (Wixted, 2004). Also in line with this account was the observation that novel words formed from the least useful size relationships established across the B-A-C word triplets were most vulnerable to waking interference on the generalisation measure. However, an active account of consolidation was supported through the sleep-specific resurgence in accuracy for A-B list size relationships (Drosopoulos, Windau, et al., 2007). Interestingly, this spontaneous recovery was observed for both congruent and incongruent list memories, despite the chance levels of performance in the latter condition. This led to the suggestion that RI recovery may not be contingent on the generalised recuperative benefits attributed to systems-level consolidation. Instead, given the lack of expressed relationships between sleep physiology and the behavioural measures, it was concluded that other unmeasured components of the sleep state may be contributing to the adaptive phenomenon of RI recovery (Fenn et al., 2003).

6.2 The influence of prior knowledge on memory consolidation

All of the investigations in this thesis, excluding Experiment 2, manipulated the congruency of novel information in relation to prior knowledge. In order to establish relevant behavioural patterns resulting from this manipulation Table 40 provides a summary of the key findings and, where relevant, the impact of the congruency manipulation in these experiments. Note that the between-participant time delays in each investigation (in the left hand column) are defined as Time 1 (T1) and Time 2 (T2) respectively.

6.2.1 Overview

A common feature of Experiments 1-5 was the implementation of modified paradigms with an established presence in the memory consolidation literature. The previous iterations of these paradigms featured stimuli that were either entirely novel (Ellenbogen et al., 2007) or carefully controlled to preclude any meaningful semantic relationships (Drosopoulos, Schulze, et al., 2007). Experiments 1-5 adapted these paradigms to feature novel stimuli that related directly to pre-existing semantic memory.

As an overview, Table 40 demonstrates that this modification is sufficient to bring about multiple time-associated (Experiments 1, 3 and 4) and sleep-associated (Experiment 5) changes to behaviour. Throughout this thesis many of these changes have, with varying degrees of certainty, have been attributed to the influences of memory consolidation. In particular, the speeding of novel word TIs (Experiment 1), and the spontaneous recovery of A-B list memories after sleep (Experiment 5), show that inter-stimuli relationships alone are sufficient to induce consolidation-associated changes to behaviour only previously observed using neutral stimuli.

At their broadest level many of the findings in Table 40 help further delineate the somewhat variable time course attributed to consolidation. In Chapter 1 it was noted that, depending on the methodology used, the influences of consolidation can be seen over months and years (MacKinnon & Squire, 1989; Ryan et al., 2001) days and weeks (Bontempi et al., 1999; Tse et al., 2007) and in some cases, just one day after training (Takashima et al., 2009; Tamminen et al., 2010). Several of the measures in the present thesis indicate that when recently acquired memories relate to prior knowledge consolidation can be seen within the latter, relatively brisk, time course. While this does not preclude additional integration over successive nights (Meeter & Murre, 2004; Walker, Brakefield, Allan Hobson, et al., 2003) these data nevertheless support previous observations indicating that memories that share a relationship with prior knowledge can be consolidated within the first few hours after learning (van Kesteren, Fernández, et al., 2010; van Kesteren, Rijpkema, et al., 2010).

6.2.2 The impact of congruency

Despite the numerous key findings presented in Table 40 (see ‘Difference at T2’) it is also noteworthy that the vast majority of these effects remained undifferentiated by the congruency manipulation itself. This commonality, shared by many of the central observations made in this thesis, is particularly striking given the marked impact of semantic incongruency on the underlying levels of accuracy for these measures. In many of these cases incongruent memory can be seen to diminish to chance (~50%) or below chance levels (<50%) after a delay (the three columns to the right of Table 40). If prior knowledge can facilitate the swift integration and recovery of new memories, why are these effects not influenced by the congruency of this information (Tse et al., 2007; van Kesteren, Fernández, et al., 2010; van Kesteren, Rijpkema, et al., 2010)? This apparent discrepancy can be reconciled on multiple levels.

Firstly, it was noted in Experiment 1 that incongruent stimuli relationships had been reinterpreted by participants prior to a consolidation opportunity. This can be seen in the far left hand column in Table 40, where below chance levels of accuracy are evident for incongruent TIs at both the 20 minute (T1) and 24 hour (T2) time points. In this particular experiment incongruent information was introduced covertly in training, with no instructions provided on how to deal with these conflicting associations when they arose. Consequently, it would appear that in order to make sense of the conflicting semantic information participants applied their own expectations of size to these relationships during, or shortly after, training. As a result diverging levels of accuracy can be seen in parallel with integration effects across 24 hours (T2). Consequently, Experiment 1 demonstrates not only that prior knowledge can facilitate mnemonic integration, but that existing memory may also be implemented to compensate for a degree of environmental noise (i.e. incongruency) during training (Reyna & Brainerd, 1990).

Secondly, with regards to the key findings derived from the A-B, A-C list learning paradigm featured in Experiments 3-5, a different set of factors may be in play. Namely, in these cases the benefits of consolidation could not be statistically distinguished from the absolute drop in accuracy levels seen in the incongruent condition. In Experiments 3 and 4 the interaction between time delay and list retention was induced through a reduction in interlist RI across time, an effect which occurred in parallel with an absolute reduction in the capacity to remember incongruent size relationships. Moreover, in Experiment 5, the incongruent list memories ‘recovered’ to chance levels, similarly forming part of the interaction driving this effect. While the influences of consolidation-associated recovery could not be fully disentangled across the congruency manipulation, in all three experiments incongruent list memories had similarly dropped to chance levels at T2. As such, congruency clearly had an effect on these measures, and indeed this outcome can be considered adaptive as only congruent (i.e. useful) size relationships were maintained at functional levels across time.

A third factor to consider, originally alluded to in Chapter 2, relates to potential crossover effects over the congruency manipulation. Previous studies exploring the role of congruency on memory consolidation are typically designed to preclude an influence of one congruency condition on the other. This can be seen in experiments that implement congruent and incongruent conditions through between-participant designs (van Kesteren, Fernández, et al., 2010), separate training contexts (Tse et al., 2007) or item-specific semantic knowledge (van Kesteren, Rijpkema, et al., 2010). Unlike these investigations, the congruency conditions in the present thesis were trained in parallel meaning that congruent size relationships may have provided useful cues regarding the treatment of incongruent information. It remains difficult to determine the full extent that this factor may play in some of the parallel changes to behaviour seen across this manipulation. Nevertheless, a consolidation-opportunity may have compounded any crossover effects, particularly given that sleep has been found to facilitate both the abstraction and application of implicitly learned rules to novel information (Durrant et al., 2011; Gómez et al., 2006; Wagner et al., 2004).

Table 40. Summary of the key findings and impact of the congruency manipulation across Experiments 1-5

		Memory change across time		Congruency accuracy levels		
Experiment	Measure	Difference at T2	Congruency interaction?	Main effect of Congruency	Congruent accuracy	Incongruent accuracy
Experiment 1 T1 = 20 minutes T2 = 24 hours	*Central premises*	Faster RTs (*)	No	Con > Incon (*)	T1 >50% T2 > 50%	T1 >50% T2 ~50%
	Novel word TIs	Faster RTs (**)	No	Con > Incon (***)	T1 >50% T2 > 50%	T1 >50% T2 > 50%
	NW size generalisations	*None*	No	No	N/A	N/A
Experiment 2 T1 = 12 hours sleep T2 = 12 hours wake	*Central premises*	*None*	-	-	T1 >50% T2 > 50%	-
	Novel word TIs	*None*	-	-	T1 >50% T2 > 50%	-
	NW size generalisations	*None*	-	-	T1 >50% T2 > 50%	-
Experiment 3 T1 = 20 minutes T2 = 24 hours	*Correct list recall %*	Lower recall % (*)	No	No	N/A	N/A
	Incorrect list recall %	*None*	No	No	N/A	N/A
	List size judgements	Lower A-C % (**)	No	Con > Incon (***)	T1 >50% T2 > 50%	T1 >50% T2 ~50%
	NW size generalisations	B-A-C size focus (***)	Yes	Con < Incon (**)	T1 >50% T2 > 50%	T1 >50% T2 > 50%
Experiment 4 T1 = 20 minutes T2 = 24 hours	*Correct list recall %*	Lower recall % (***)	No	No	N/A	N/A
	Incorrect list recall %	Higher recall (**)	No	No	N/A	N/A
	List size judgements	Lower A-C % (**)	No	Con > Incon (***)	T1 >50% T2 > 50%	T1>50% T2 ~50%
	NW size generalisations	*None*	No	Con > Incon (*)	T1 >50% T2 > 50%	T1 >50% T2 > 50%
Experiment 5 T1 = 12 hours sleep T2 = 12 hours wake	*Correct list recall %*	Higher recall % (*)	No	No	N/A	N/A
	Incorrect list recall %	Lower recall (**)	No	No	N/A	N/A
	List size judgements	Higher A-B % (*)	No	Con > Incon (***)	T1 ~50% T2 >50%	T1<50% T2~50%
	NW size generalisations	No B-A-C size focus (***)	No	No	T1 >50% T2 > 50%	T1 >50% T2 > 50%

Notes: Values in parentheses represent the significance value of a given finding (*) indicates *p*<.05, (**) indicates *p*<.01, and (***) indicates *p<.001*. T1 and T2 indicate the two compared time points in a given experiment. Levels of congruent and incongruent accuracy reflect the relationships between accuracy performance and chance (~50%) as determined through one-sample t-tests. Note that Experiment 2 summarises the primary 12:12 sleep/wake conditions only.

6.2.3 Specific contributions to the existing literature

At a more specific level several the experiments presented in this thesis provide several novel contributions to the existing literature, particularly with regards to impact of consolidation on TI performance and the phenomenon of retroactive interference. The nature of these contributions will be briefly outlined below.

6.3.4.1 Divergent routes of consolidation

In Chapter 2 it was argued that the consolidation-associated enhancements in TI performance may differ when semantic information is included in the training of premises. The research exploring the influences of consolidation using this paradigm thus far have consistently uncovered a selective enhancement in the accuracy levels of 'B>E' TIs (Ellenbogen et al., 2007). These inferences are derived from the strongly reinforced 'A>' and '<F' premises at the ends of the schematic hierarchy (the dashed boxes in Figure 34). The consolidation-associated enhancement of these particular TIs does not occur when participants become explicitly aware of the overarching stimuli relationships (e.g. 'A>B>C>D>E>F') or when premises are trained through passive observation, as opposed to associative reinforcement (Werchan & Gómez, 2013). It was argued in Chapter 2 that this pattern of integration is reflective of procedural, rather than declarative memory consolidation (Dudai, 2004). Indeed the literature specific to the field of TI indicates that associative strength of premises alone can support the TI capabilities in animals (McGonigle & Chalmers, 1977; von Fersen et al., 1991) as well as 'unaware' human participants (Frank et al., 2006, 2005).

Experiment 1, summarised on the right of Figure 34, demonstrates for the first time that the pattern of consolidation-associated integration seen in the TI paradigm can differ when prior knowledge is engaged during training. Here, levels of accuracy performance remained equivalent across time for B>E inferences, while integration was simultaneously evident across the central elements of the schematic dimension (the solid black boxes in Figure 34). It is worth noting that the B>D inferences in Experiment 1, as with previous investigations, required a comparison between two elements of entirely new information (textures or novel words). The differences in integration between these experiments therefore appears to be contingent on the presence or absence of semantic information during training. In Chapter 2 it was speculated that consolidation may diverge down two different routes of integration (procedural or declarative) in order to optimise memory for future use under these different circumstances (Wagner et al., 2004).

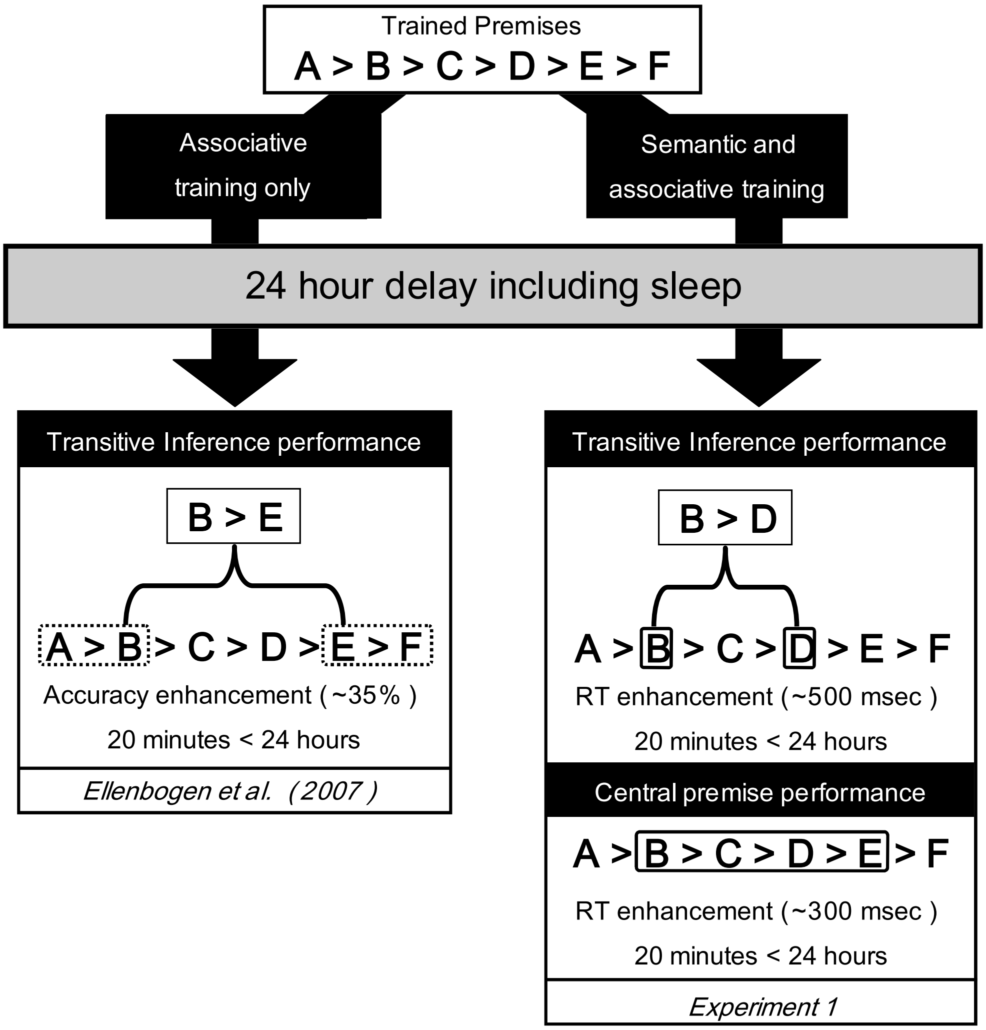


Figure 34. Diverging patterns of consolidation-associated integration seen in Experiment 1 (right) relative to previous observations made using the transitive inference paradigm (left).

6.3.4.1 The time course of spontaneous recovery

As established in Chapters 4 and 5 the observation that recently established semantic associations can spontaneously recovery from RI, to the author’s knowledge, represents a novel finding in this thesis. However, despite featuring equivalent training protocols the complementary time points obtained in Experiment 3 (20 minutes and 24 hours) and Experiment 5 (12 hours of sleep or wakefulness) have yet to be considered together. In order to clarify the both the time course of spontaneous recovery, and the active role of consolidation in these effects, Experiments 3 and 5 are therefore briefly considered together here. This comparison, presented in Table 41, focuses on congruent list memory given the otherwise chance levels of performance seen in the incongruent condition. The statistical analysis for this figure can be found in Appendix 20.

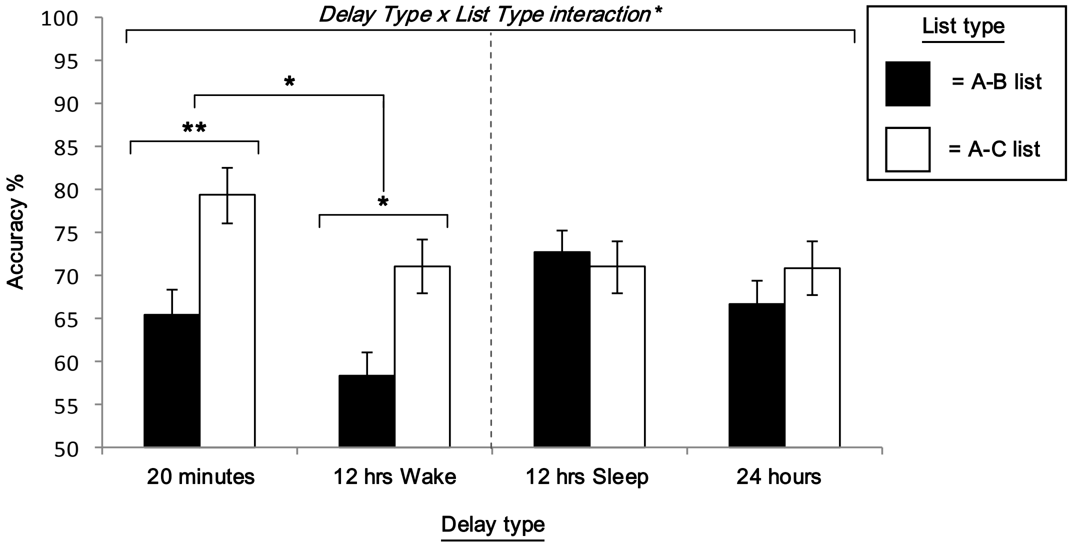


Figure 35. The role of consolidation in the spontaneous recovery and stabilisation of congruent list memories over 24 hours. Significance levels (*) indicates *p*<.05, and (**) indicates *p*<.01. Error bars represent the SE. The grey dashed line indicates a consolidation opportunity.

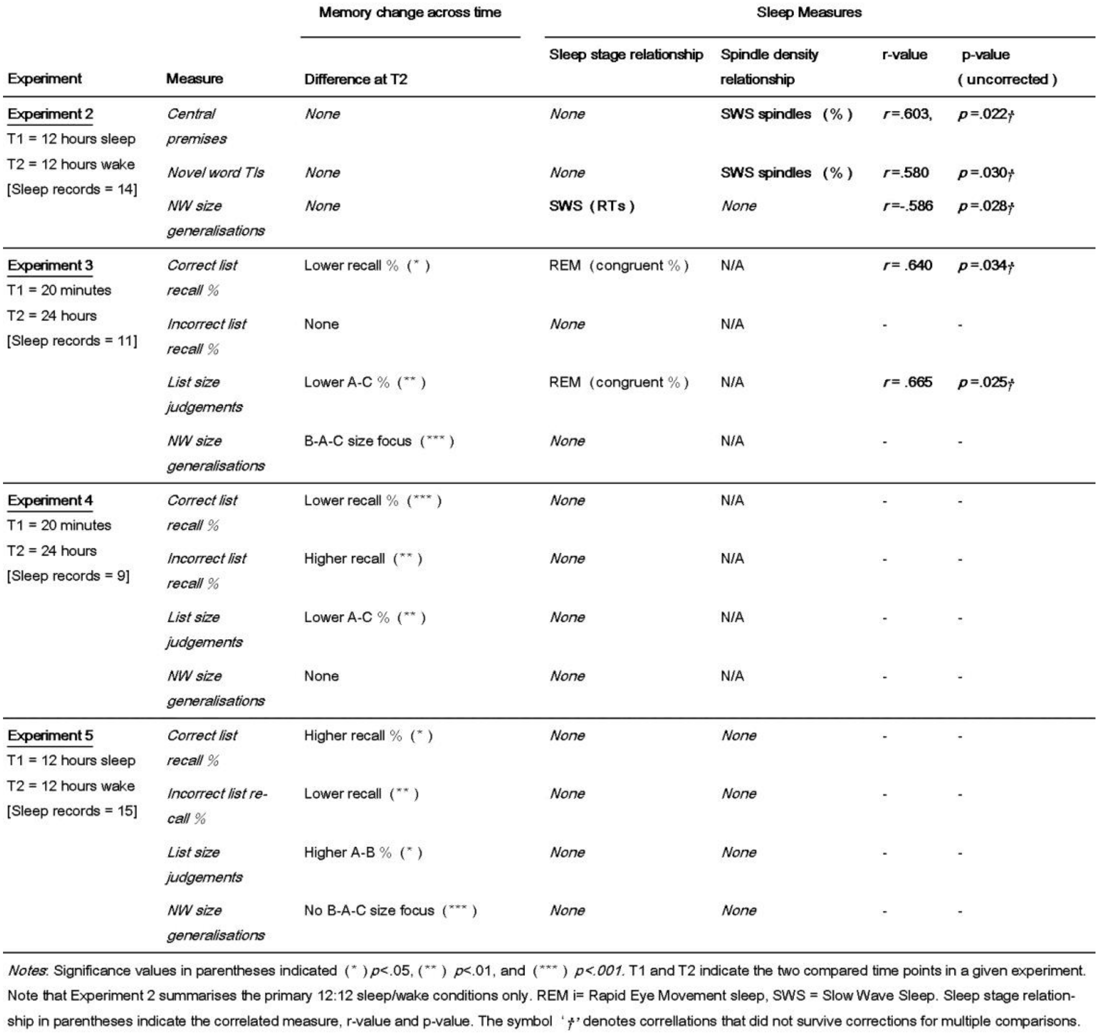
The four time points presented in Figure 35 serve to delineate the involvement of consolidation on congruent list memories over 24 hours. Prior to a consolidation opportunity (the grey dashed line) RI effects are clearly evident between the A-B and A-C list memories, as exemplified in the relatively low levels of A-B list accuracy after 20 minutes and 12 hours of wakefulness. Only after a period of sleep do A-B list memories recover from RI, an effect that continues to remain present at the later 24 hour time point. These data clearly support the interpretation made in Chapter 5 that sleep is beneficial to the recovery of recently acquired semantic memories (Drosopoulos, Schulze, et al., 2007; Ekstrand, 1967).

Interestingly, Figure 35 provides a second insight into the time course of consolidation not originally evident when Experiments 3 and 5 were considered separately. Prior to a period of sleep, memory for both the A-B and A-C lists can be seen to deteriorate over wakefulness, as seen in the drop in accuracy across 12 hours of wakefulness. However, after sleep there is no further decay in the memory of B-A-C word triplets, as seen in the equivalent levels of performance at the later 24 hour time point. This indicates that, as well as facilitating spontaneous recovery, a consolidation-opportunity can help stabilise memories by protecting them from subsequent interference (Walker & Stickgold, 2004; Wixted, 2004). It has been previously shown that A-B list memories will become more resistant to RI if the secondary A-C list is acquired after a period of sleep but not wakefulness (Ellenbogen et al., 2009; Ellenbogen, Hulbert, et al., 2006). The results in Figure 35 expand on these observations by showing that learning *both* A-B and A-C lists prior to a consolidation opportunity can bolster their resistance to the more general effects of RI across wakefulness.

6.4 The sleep associated mechanisms underlying consolidation

A second aim of this thesis was to elucidate the potential involvement of sleep with regards to the behavioural changes noted in the previous section. The majority of the experiments contained in this thesis featured some kind of measure relating to sleep physiology, which were then correlated with memory performance after a night of sleep. The resulting relationships are summarised in Table 41. Note that only Experiments 2 and 3 are included in this table given the null results of the sleep measures in Experiments 4 and 5.

Table 41. Behavioural measures and their uncorrected relationship with sleep physiology in Experiments 2-5.



		Memory change across time	Sleep Measures			
Experiment	Measure	Difference at T2	Sleep stage relationship	Spindle density relationship	r-value	p-value (uncorrected)
Experiment 2 T1 = 12 hours sleep T2 = 12 hours wake [Sleep records = 14]	*Central premises*	*None*	*None*	SWS spindles (%)	r=.603,	p=.022†
	Novel word TIs	*None*	*None*	SWS spindles (%)	r=.580	p=.030†
	NW size generalisations	*None*	SWS (RTs)	*None*	r=-.586	p=.028†
Experiment 3 T1 = 20 minutes T2 = 24 hours [Sleep records = 11]	*Correct list recall %*	Lower recall % (*)	REM (congruent %)	N/A	r= .640	p=.034†
	Incorrect list recall %	None	*None*	N/A	-	-
	List size judgements	Lower A-C % (**)	REM (congruent %)	N/A	r= .665	p=.025†
	NW size generalisations	B-A-C size focus (***)	*None*	N/A	-	-
Experiment 4 T1 = 20 minutes T2 = 24 hours [Sleep records = 9]	*Correct list recall %*	Lower recall % (***)	*None*	N/A	-	-
	Incorrect list recall %	Higher recall (**)	*None*	N/A	-	-
	List size judgements	Lower A-C % (**)	*None*	N/A	-	-
	NW size generalisations	None	*None*	N/A	-	-
Experiment 5 T1 = 12 hours sleep T2 = 12 hours wake [Sleep records = 15]	*Correct list recall %*	Higher recall % (*)	*None*	*None*	-	-
	Incorrect list recall %	Lower recall (**)	*None*	*None*	-	-
	List size judgements	Higher A-B % (*)	*None*	*None*	-	-
	NW size generalisations	No B-A-C size focus (***)	*None*	*None*	-	-

Notes: Significance values in parentheses indicated (*) $p<.05$, (**) $p<.01$, and (***) $p<.001$. T1 and T2 indicate the two compared time points in a given experiment. Note that Experiment 2 summarises the primary 12:12 sleep/wake conditions only. REM i= Rapid Eye Movement sleep, SWS = Slow Wave Sleep. Sleep stage relationship in parentheses indicate the correlated measure, r-value and p-value. The symbol '†' denotes correllations that did not survive corrections for multiple comparisons.

6.4.1 Overview

Despite the key behavioural findings outlined in the previous sections the sleep measures obtained this thesis did not reveal consistent findings across these investigations. Only two of the four experiments featuring sleep measures revealed any meaningful association with sleep activity, and in all cases these expressed relationships did not survive correction for multiple comparisons (Table 41). Moreover, the null behavioural findings in

Experiment 2, and extensive participant dropout in Experiment 3, further limit the extent that these findings can generalised.

However, despite these caveats, all five of the physiological relationships displayed in Table 41 are aligned in a direction suggestive of a positive contribution of sleep architecture on memory retention. If these associations were purely an artefact of multiple comparisons we might expect a mixture of positive and negative associations between sleep activity and memory performance. Moreover, each of these findings has some empirical grounding regarding the predictive nature of these relationships on memory.

6.4.2 Potential contributions of sleep activity on memory consolidation

In Experiment 2 weak relationships were positively expressed between slow wave sleep (SWS) spindle activity and the accuracy levels of both novel word TIs and central premise items. As noted throughout this thesis sleep spindles, particularly when aligned with this sleep stage, form a key mechanistic component of the standard model of memory consolidation (Mölle et al., 2009; Siapas & Wilson, 1998; Steriade, 2001). Moreover, in Experiment 1 it was shown that central premise and novel word TIs displayed the greatest performance enhancement across 24 hours, suggesting that these measures are the most susceptible to consolidation-associated integration. Indeed, the pattern of integration seen in Experiment 1, as highlighted in Figure 35, strongly suggests that new information was enhanced over time through its association with prior knowledge, an effect that was potentially mediated by systems-level consolidation during sleep (Frankland & Bontempi, 2005; McClelland et al., 1995).

When Experiments 1 and 2 are considered together, the correlations with sleep physiology seen in the latter investigation may represent weakened forms of what otherwise might have been a considerable influence of sleep on memory integration. These effects may have been circumvented in Experiment 2 through the through the over priming of congruent information (Stickgold, 2009), a factor potentially reflected in the lower number of sleep spindles relative to other investigations featuring the same spindle detection algorithm (cf. Schabus et al., 2006; Tamminen, 2010). As such, although it cannot be said from these data alone that systems-level consolidation supported the integration seen in Experiment 1, they do form the basis of a testable hypothesis suggesting that this was indeed the case. The next logical step would be to implement the same paradigm featured in Experiment 1 in a 12:12 sleep/wake design to test this prediction.

Lastly, the implication that rapid eye movement (REM) sleep may support mnemonic recovery, as suggested in Experiment 3 (Table 41) is generally less viable than the potential contribution of SWS on integration noted above (Diekelmann et al., 2009). In particular, this relationship was not replicated in Experiment 5, which featured higher quality electrophysiological recordings, more participants (n=15) and a direct measure of spontaneous recovery (Figure 35). That said, the relationship with REM sleep in Experiment 3 did align with one model of mnemonic recovery which, to date, has not received much behavioural support (Norman et al., 2005). In particular this model proposes that it is theta-wave oscillations in REM sleep, rather that the REM sleep stage itself, that support the recovery of A-B associations during sleep. On this basis it was suggested in Chapter 5 that future investigations would benefit from a measure of theta-wave activity across the entire sleep state to help elucidate the as-yet unspecified mechanistic contributions of sleep on mnemonic recovery (Schabus et al., 2005).

6.5 Weaknesses of the present thesis

6.5.1 Circadian confounds

The measures from both Experiments 2 and 5 indicated that some form of circadian confound may have been present across the time delay manipulation. In Experiment 2 in particular, there was a slowing in the RTs on some of the baseline measures such as real word comparisons, suggestive of time of day effects (Tilley & Warren, 1983). It remains difficult to determine the extent that this factor may have influenced the critical measures of integration Experiment 2, particularly given the null behavioural findings in that experiment. In Experiment 5 there was also evidence that participants were less alert after a period of sleep relative to wakefulness on both subjective and objective baseline measures. However, in this case these differences were accounted for in analysis, and were shown to have no effect on the principle behavioural findings.

6.5.2 The issue of multiple comparisons

In Chapter 2 the issue of multiple comparisons between sleep physiology and behaviour was raised. Namely, there is currently no standardised approach in the sleep literature with regards to multiple measures of sleep physiology and behaviour. This is an important issue to consider given both the multifaceted nature of the sleep state (Conte & Ficca, 2012; Stickgold, 2005) and the rich datasets obtained through the exploratory investigations presented in this thesis. As a result, a large number of potential correlations could have potentially been performed, bringing with them an increased risk of a type I statistical error through multiple comparisons. On the other hand, extensive corrections for multiple comparisons might have masked genuine relationships between consolidation and sleep physiology, thus contributing to a type II statistical error.

There was no easy way around this problem however attempts were made to address this throughout this thesis. Where possible, behavioural factors were reduced down to composite measures reflecting the principle effects of interest in order to limit the number of derived correlations. Moreover, a consistent rule was applied through this thesis regarding the treatment of multiple comparisons (first outlined in Chapter 3). In the absence of firm guidelines in the existing literature this approach was intended to strike a balance between potential type I and type II errors, however it remains hard to say if the resulting relationships were adjusted too conservatively or liberally.

Future studies may benefit from the use of a false discovery rate (FDR) to correct for multiple comparisons, a less conservative adjustment than the bonferroni corrections implemented in this thesis. Although FDRs are not typically found in the sleep literature they have become the standard method of adjustment in other areas of the biological sciences, and offer superior power when compared to bonferroni-type procedures (Benjamini & Hochberg, 1995).

A final point regarding the largely null findings relating to the PSG correlations presented in this thesis relates to the appropriate power levels required to register an effect. In Appendix 21 power calculations conducted across Experiments 2-5 indicate that experimental power for the PSG data was moderately low in Experiment 2 and poor in Experiments 3-5. In some ways the issues regarding power were difficult to avoid, for example (i) the known drop-out issues regarding the Zeo device, and (ii) budget and time contraints for data collection. Although the numbers of participants chosen for these experiments was based on previous research, future studies would perhaps be best served by calculating the appropriate number of participants required for adequate levels of experimental power prior to being conducted.

6.5.3 Other Limitations

A noted weakness in the methodology of all five experiments contained in this thesis was the implementation of the Stanford Sleepiness Scale (SSS) in lieu of the generally preferred (and more valid) Karloniska Sleepiness Scale (KSS). The use of the SSS, in conjunction with a visual analogue scale, was originally deemed to be a suitable measure of sleepiness based on the advice of other sleep researchers (in particular Dr. Tamminen). Despite this the SSS remains widely used and it was only after the completion of this thesis was the KSS discovered. As such, future studies based on the work contained in this thesis may wish to implement the KSS and not the SSS to gauge circadian influences.

Another important issue relates to the internal validity of the sleep scoring measures in this thesis. Firstly, the Zeo sleep devices (Experiments 3 and 4) and polysomnography (PSG) systems (Experiments 2 and 5) are two entirely different instruments for recording sleep with unclear equivalency (Griessenberger et al., 2012) . Secondly, different sleep scoring systems were used for the PSG records in Experiments 2 and 5. Namely, the R&K (Rechtschaffen & Kales, 1968) scoring system was found to have several limitations (Himanen & Hasan, 2000) hence the preference for the more modern AASM system in Experiment 5. Lastly, my own skill in sleep scoring almost certainly improved across the 4 years it took to conduct these experiments. For all of these reasons, and despite the precautions taken to cross score the sleep measures, the consistency of the sleep records data may not be sufficiently consistent.

Lastly, although the protocols for Experiments 1-5 were developed from numerous published studies, problems regarding circadian influences may still remain. For example, the detrimental influence of sleep inertia was previously thought to be circumvented by testing individuals 30 minutes after arousal as opposed to immediately (See Tamminen et al, 2010). However, recent work suggests that the effects of sleep inertia may still be present up to three hours after arousal (Groeger, Lo, Burns, & Dijk, 2011). As such future studies may wish to extend testing beyond the 30 minute delay featured in Experiments 2 and 5.

6.6 Conclusions and future work

6.6.1 Main contributions of this thesis

In the sections above the key findings of this thesis have been outlined with regards to the impact of prior knowledge on memory consolidation. These observations are drawn together in Figure 36 which summarises the inferred time course and influence of consolidation on memory. The presence of semantic information in training was found to support the numerous changes to memory (A-E) present in Figure 36 which, to date, have not been demonstrated using the TI or A-B, A-C list learning paradigms.

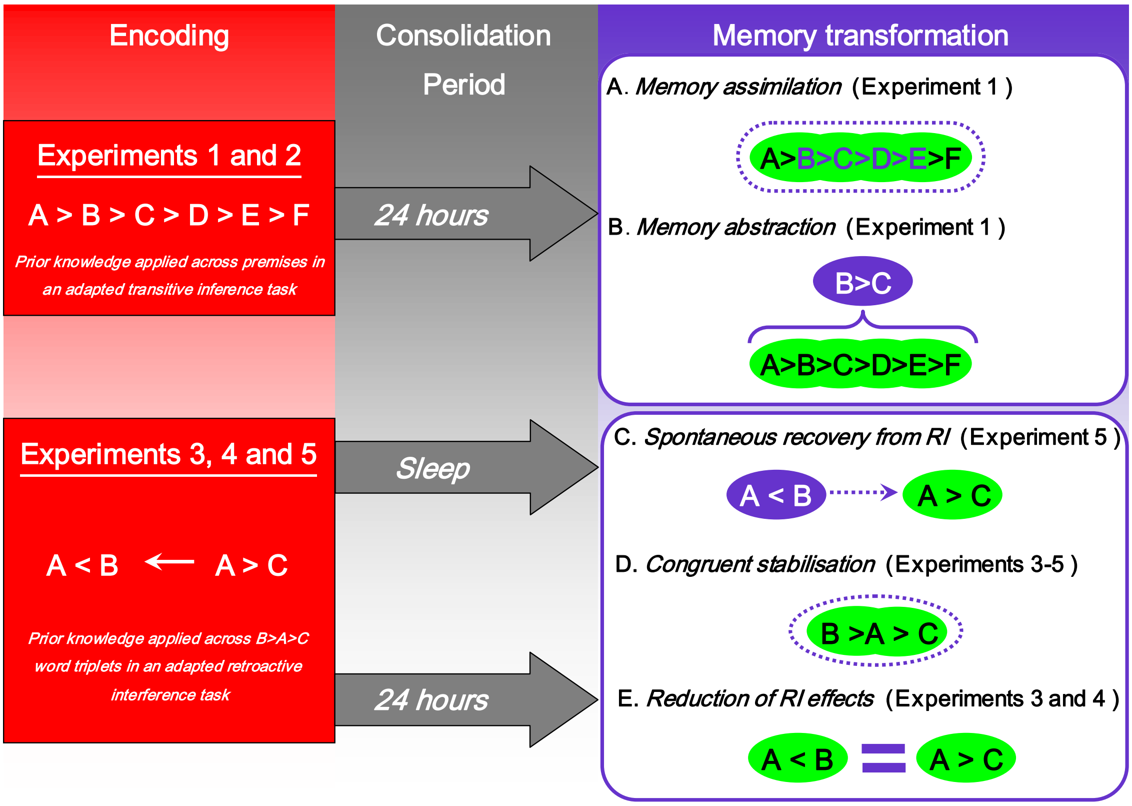


Figure 36. The main behavioural findings of the present thesis. Prior knowledge has been shown to promote the following consolidation-associated effects: (A) memory assimilation, (B) memory abstraction, (C) the spontaneous recovery of memories from retroactive interference (RI), (D) the stabilisation of memories, and (E) the reduction of RI effects over time. The combined influence of consolidation and prior knowledge are highlighted in purple in this figure.

The impact of semantic congruency differed between the two approaches used to explore the principle research question in this thesis (the red boxes in Figure 36). In Experiment 1 it was shown that incongruent information was interpreted in a congruent manner shortly after training, resulting in parallel integration across the congruency manipulation (Figure 36, A and B). In Experiments 3-5 the retention of incongruent list memories largely dropped to chance levels after a consolidation-opportunity. As such, the influences of consolidation shown in Experiments 3-5 (Figure 36, C-E) only appear benefit the future application of semantically congruent information. Taken together these findings make several contributions to our current understanding of how prior knowledge may influence the consolidation of new information, which can be summarised in the following three points.

1. Semantic information can contribute to the rapid assimilation and abstraction of recently formed memories over a 24 hour period. Moreover, this factor has been shown to facilitate the integration of both congruent and incongruent information over time, highlighting the role of prior knowledge in both the encoding and post-encoding consolidation of new memories.

2. It was demonstrated using a modified TI paradigm that consolidation-associated integration may differ depending on the presence, or absence, of semantic information during training. This observation suggests that consolidation will optimise memory in different ways depending on the applicability of new information to existing declarative memory.

3. Semantic information can facilitate both the consolidation-associated recovery and stabilisation of memories. These effects in combination have been shown to be particularly beneficial in counteracting the detrimental influences of RI for congruent, but not incongruent, memories after a period of sleep.

6.6.2 Future work

The conclusions outlined above raise a number of questions for future study. Firstly, while systems-level integration was strongly implicated in the findings of Experiment 1, particularly given the provisional association with SWS spindle activity seen in Experiment 2, this form of consolidation can only be inferred from the present data. As TI performance is strongly associated with hippocampal activation shortly after training (Greene, Gross, Elsinger, & Rao, 2006; Heckers et al., 2004) this paradigm appears to be well suited to explore the neurological shifts in retrieval dependency attributed to systems-level consolidation (Frankland & Bontempi, 2005). For example, it might be predicted that TI performance will manifest a greater degree of functional connectivity between the hippocampus and medial prefrontal cortex after a consolidation opportunity (Frankland & Bontempi, 2006; van Kesteren et al., 2012).

Furthermore, while the modified TI and RI paradigms featured in this thesis show promise, additional research would benefit from establishing how generlalisable these findings are. For example, no integration was observed in Experiment 2 relative to Experiment 1 once incongruent condition was removed from training. It therefore remains unclear if a degree of incongruency is both sufficient and necessary for the consolidation-associated changes to arise in Figure 36. In particular, would spontaneous recovery still emerge after a period of consolidation if all of the semantic relationships were trained as congruent?

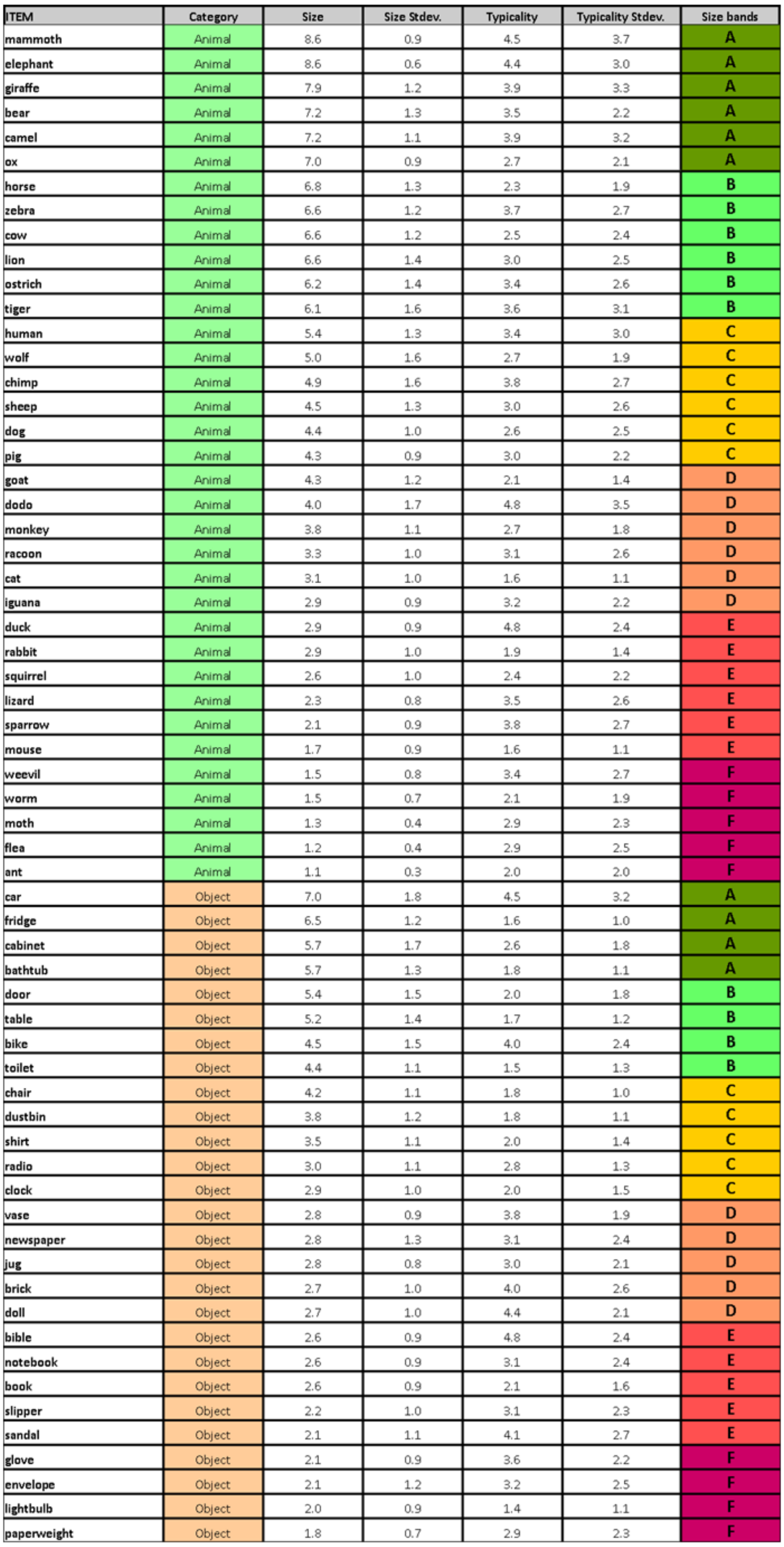
Lastly, as the experiments in this thesis demonstrate that semantic information is sufficient to bring about consolidation-associated changes to behaviour there is considerable room to build on these findings. As a starting point, a relevant question is whether novel word integration on these tasks will manifest differently depending on type of semantic information featured in training. For example, animal items may have more underlying features in common (i.e. they are living, reproduce, etc.) than household objects, which share a more eclectic range of properties (Murphy, 2004). As such, animal words may confer a more salient set of semantic properties amenable to rapid semantic integration. Recent research has shown that novel word integration can indeed be influenced by the semantic properties of affiliated real word items, such semantic neighbourhood density (Tamminen, Lambon Ralph, et al., 2012). The time therefore appears to be right to explore how such factors may influence mnemonic integration on more complex tasks, such as those featured in this thesis.

Appendices

Appendix 1

Size band allocation and ratings (n=20) for Experiments 1 and 2.

Ratings for land animals (n=37) and household objects (n=37) were judged in relation to each other using a 1-9 Likert scale. Hierarchy construction was derived from protocols outlined by Rubinsten, Henik, Berger, and Shahar-Shalev (2002) with the removal of conceptually 'atypical' items (>5) leaving 35 land animals and 27 household items for size band allocation.

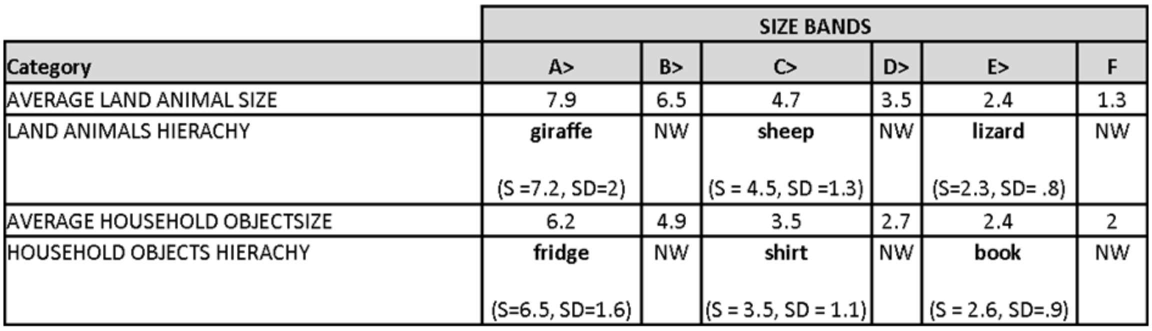


ITEM	Category	Size	Size Stdev.	Typicality	Typicality Stdev.	Size bands
mammoth	Animal	8.6	0.9	4.5	3.7	A
elephant	Animal	8.6	0.6	4.4	3.0	A
giraffe	Animal	7.9	1.2	3.9	3.3	A
bear	Animal	7.2	1.3	3.5	2.2	A
camel	Animal	7.2	1.1	3.9	3.2	A
ox	Animal	7.0	0.9	2.7	2.1	A
horse	Animal	6.8	1.3	2.3	1.9	B
zebra	Animal	6.6	1.2	3.7	2.7	B
cow	Animal	6.6	1.2	2.5	2.4	B
lion	Animal	6.6	1.4	3.0	2.5	B
ostrich	Animal	6.2	1.4	3.4	2.6	B
tiger	Animal	6.1	1.6	3.6	3.1	B
human	Animal	5.4	1.3	3.4	3.0	C
wolf	Animal	5.0	1.6	2.7	1.9	C
chimp	Animal	4.9	1.6	3.8	2.7	C
sheep	Animal	4.5	1.3	3.0	2.6	C
dog	Animal	4.4	1.0	2.6	2.5	C
pig	Animal	4.3	0.9	3.0	2.2	C
goat	Animal	4.3	1.2	2.1	1.4	D
dodo	Animal	4.0	1.7	4.8	3.5	D
monkey	Animal	3.8	1.1	2.7	1.8	D
racoon	Animal	3.3	1.0	3.1	2.6	D
cat	Animal	3.1	1.0	1.6	1.1	D
iguana	Animal	2.9	0.9	3.2	2.2	D
duck	Animal	2.9	0.9	4.8	2.4	E
rabbit	Animal	2.9	1.0	1.9	1.4	E
squirrel	Animal	2.6	1.0	2.4	2.2	E
lizard	Animal	2.3	0.8	3.5	2.6	E
sparrow	Animal	2.1	0.9	3.8	2.7	E
mouse	Animal	1.7	0.9	1.6	1.1	E
weevil	Animal	1.5	0.8	3.4	2.7	F
worm	Animal	1.5	0.7	2.1	1.9	F
moth	Animal	1.3	0.4	2.9	2.3	F
flea	Animal	1.2	0.4	2.9	2.5	F
ant	Animal	1.1	0.3	2.0	2.0	F
car	Object	7.0	1.8	4.5	3.2	A
fridge	Object	6.5	1.2	1.6	1.0	A
cabinet	Object	5.7	1.7	2.6	1.8	A
bathtub	Object	5.7	1.3	1.8	1.1	A
door	Object	5.4	1.5	2.0	1.8	B
table	Object	5.2	1.4	1.7	1.2	B
bike	Object	4.5	1.5	4.0	2.4	B
toilet	Object	4.4	1.1	1.5	1.3	B
chair	Object	4.2	1.1	1.8	1.0	C
dustbin	Object	3.8	1.2	1.8	1.1	C
shirt	Object	3.5	1.1	2.0	1.4	C
radio	Object	3.0	1.1	2.8	1.3	C
clock	Object	2.9	1.0	2.0	1.5	C
vase	Object	2.8	0.9	3.8	1.9	D
newspaper	Object	2.8	1.3	3.1	2.4	D
jug	Object	2.8	0.8	3.0	2.1	D
brick	Object	2.7	1.0	4.0	2.6	D
doll	Object	2.7	1.0	4.4	2.1	D
bible	Object	2.6	0.9	4.8	2.4	E
notebook	Object	2.6	0.9	3.1	2.4	E
book	Object	2.6	0.9	2.1	1.6	E
slipper	Object	2.2	1.0	3.1	2.3	E
sandal	Object	2.1	1.1	4.1	2.7	E
glove	Object	2.1	0.9	3.6	2.2	F
envelope	Object	2.1	1.2	3.2	2.5	F
lightbulb	Object	2.0	0.9	1.4	1.1	F
paperweight	Object	1.8	0.7	2.9	2.3	F

Appendix 2

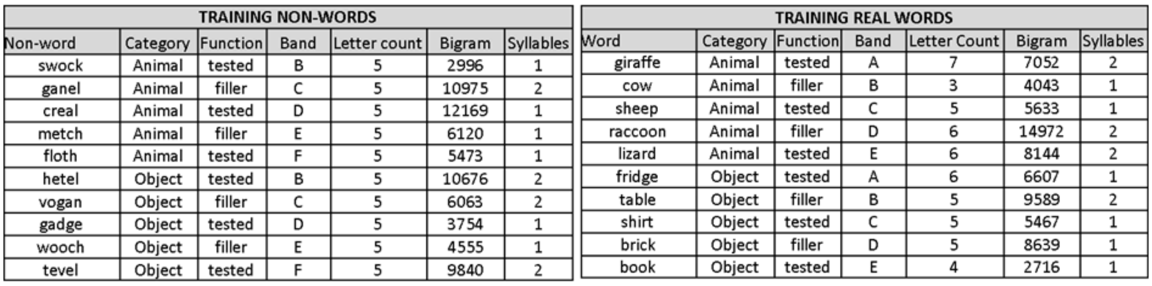
Additional information regarding size band allocation and stimuli organisation for Experiments 1 and 2.

TABLE 1. Stimuli derived from the size bands in Appendix 1.



	SIZE BANDS					
Category	**A>**	**B>**	**C>**	**D>**	**E>**	**F**
AVERAGE LAND ANIMAL SIZE	7.9	6.5	4.7	3.5	2.4	1.3
LAND ANIMALS HIERACHY	**giraffe** (S =7.2, SD=2)	NW	**sheep** (S = 4.5, SD =1.3)	NW	**lizard** (S=2.3, SD= .8)	NW
AVERAGE HOUSEHOLD OBJECTSIZE	6.2	4.9	3.5	2.7	2.4	2
HOUSEHOLD OBJECTS HIERACHY	**fridge** (S=6.5, SD=1.6)	NW	**shirt** (S = 3.5, SD = 1.1)	NW	**book** (S = 2.6, SD=.9)	NW

TABLE 2. Bigram frequencies, word counts and syllable counts for all trained stimuli.



TRAINING NON-WORDS						
Non-word	Category	Function	Band	Letter count	Bigram	Syllables
swock	Animal	tested	B	5	2996	1
ganel	Animal	filler	C	5	10975	2
creal	Animal	tested	D	5	12169	1
metch	Animal	filler	E	5	6120	1
floth	Animal	tested	F	5	5473	1
hetel	Object	tested	B	5	10676	2
vogan	Object	filler	C	5	6063	2
gadge	Object	tested	D	5	3754	1
wooch	Object	filler	E	5	4555	1
tevel	Object	tested	F	5	9840	2

TRAINING REAL WORDS						
Word	Category	Function	Band	Letter Count	Bigram	Syllables
giraffe	Animal	tested	A	7	7052	2
cow	Animal	filler	B	3	4043	1
sheep	Animal	tested	C	5	5633	1
raccoon	Animal	filler	D	6	14972	2
lizard	Animal	tested	E	6	8144	2
fridge	Object	tested	A	6	6607	1
table	Object	filler	B	5	9589	2
shirt	Object	tested	C	5	5467	1
brick	Object	filler	D	5	8639	1
book	Object	tested	E	4	2716	1

TABLE 3. Original novel word stimuli and an size band equivalents for the new, but familiar stimuli used in the symbolic distance generalisation phase.

Category	SIZE BAND/RANK	Original	Equivalent 1	Equivalent 2
Animal	A	giraffe	bear	camel
Animal	B	swock	zebra	lion
Animal	C	sheep	dog	pig
Animal	D	creal	iguana	cat
Animal	E	lizard	sparrow	squirrel
Animal	F	floth	worm	flea
Object	A	fridge	car	cabinet
Object	B	vogan	door	bike
Object	C	shirt	dustbin	radio
Object	D	gadge	doll	jug
Object	E	book	notebook	slipper
Object	F	tevel	envelope	glove

Appendix 3

The 'explorer scenario' instructions provided to participants in Experiments 1 and 2.

TRAINING

[SCREEN 1] *'You are part of a team of explorers who have found a newly discovered area of jungle on planet earth. Having set up base camp two teams have gone out to investigate the local area while you stay at camp. Both teams radio back with new discoveries. The anthropologists have discovered a village containing a newly found tribe of humans who are using a set of items in their houses not seen before in other cultures. The biologists have also made some discoveries and have found new species of animals not seen elsewhere. Both teams have begun to catalogue these new creatures and artefacts using the local language.'*

[SCREEN 2] *'Shortly after the radio transmission there is an equipment failure at your end and you can only receive the information about the new animals and objects in pairs of words. Your task is to determine if these new things are bigger or smaller in relation to the size of animals and household objects you already know. At first you will have to guess these choices but the longer you do the task the better you will become.'*

[SCREEN 3] *'You will be asked to go through two training sessions on these new items one now and one at a later session as determined by the experimenter. The best performer on both speed and accuracy in these tasks will win a £20 Amazon gift voucher. Please do not cheat and write things down, however, as the experimenter can see you!'*

TESTING

[SCREEN 1] *'Although your research team originally planned to send you more data to catalogue from the field they have decided not to do this.*

Worried that the previous equipment failure may have affected your ability to learn the relative size of the new objects and animals they have decided to test you on what you have learned previously instead of training you on more things.'

[SCREEN 2]*'This first testing phase will look much like training phase and you will have to choose the button on your pad that corresponds to the side of the screen with the larger animal or object. However there are some changes.*

Firstly, you will no longer receive any feedback on your performance.

Secondly, you may encounter some new combinations of objects or animals you will have not seen before.

Thirdly, you now have a time limit of around ten seconds to provide an answer before a new pair of words is presented.

Please provide your best guess if you are not sure of the answer.

Please be as fast and as accurate as possible and remember there is a £20 Amazon voucher for the best performer!

Press the middle button on the pad in front of you to start this challenge.'

Appendix 4

The sleep log used in Experiments 1 and 2.

Please note the approximate time you went to bed:

Please note the approximate time of waking:

Did you have to get up during the night? Yes No

If so how many times and for how long?

....

......

....

......

Did you drink any alcohol or caffeine before you went to bed? Yes No

If you answered ‘yes’ to the above please provide more information (i.e. what it was, how much and how close it was consumed to bedtime).

....

......

....

......

Please give your best estimate of how many hours of sleep you normally get per night:

....

......

....

......

On the scale below please indicate on a scale of 1-10 how good you think your night’s sleep was (1 = bad, 10 = good).

1 2 3 4 5 6 7 8 9 10

Appendix 5

The Stanford Sleepiness Scale. (Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973)

What best describes your current level of alertness?

1. Feeling active, vital, alert, or wide awake

2. Functioning at high levels, but not at peak; able to concentrate

3. Awake, but relaxed; responsive but not fully alert

4. Somewhat foggy, let down

5. Foggy; losing interest in remaining awake; slowed down

6. Sleepy, woozy, fighting sleep; prefer to lie down

7. No longer fighting sleep, sleep onset soon; having dream-like thoughts

Appendix 6

Correlations between % of TST and various integration measures obtained in Experiment 2.

		Stage 2%	REM%	SWS%
Premises				
Acc (%)	*r*	-0.213	-0.072	0.344
	p	*0.47*	*0.81*	*0.23*
Retention (%)	*r*	-0.15	0.13	0.36
	p	*0.60*	*0.65*	*0.21*
RT (msec)	*r*	-0.10	-0.45	-0.01
	p	*0.74*	*0.11*	*0.97*
RT difference (msec)	*r*	0.01	-0.38	-0.18
	p	*0.96*	*0.19*	*0.55*
Transitive Inferences				
Non inference (%)	*r*	-0.379	-0.301	0.314
	p	0.18	0.30	0.27
Novel word inference (%)	*r*	0.009	-0.146	0.420
	p	0.98	0.62	0.13
Real word inference (%)	*r*	-0.177	-0.236	0.402
	p	0.54	0.42	0.15
Two unit inference (%)	*r*	0.116	0.268	**0.604**
	p	0.69	0.35	**0.022**†
Non inference (msec)	*r*	-0.362	0.023	0.101
	p	0.20	0.94	0.73
Novel word inference (msec)	*r*	-0.254	-0.389	-0.167
	p	0.38	0.17	0.57
Real word inference (msec)	*r*	-0.520	-0.455	0.042
	p	0.06	0.10	0.89
Two unit inference (msec)	*r*	-0.435	-0.476	-0.284
	p	0.12	0.09	0.33
Generalisation				
RW Size difference (%)	*r*	0.322	0.016	0.132
	p	*0.26*	*0.96*	*0.65*
NW Size difference (%)	*r*	-0.089	0.173	-0.136
	p	*0.76*	*0.55*	*0.64*
RW Size difference (msec)	*r*	0.512	-0.112	-0.207
	p	*0.06*	*0.70*	*0.48*
NW Size difference (msec)	*r*	0.277	0.166	-0.503
	p	*0.34*	*0.57*	*0.07*

Note: Significant correlations in bold. REM% = percent of rapid eye movement sleep in in relation to total sleep time (TST), SWS% = percent of SWS in relation to TST.† = p-values that do not survive a Bonferroni correction for multiple comparisons.

Appendix 7

Word triplet properties for Experiments 3, 4 and 5.

	Stimulus Set	B list	A novel item	C List	B list size rating	C list size Rating	Size difference (B-C)	B list word frequency (PMW)	C list word frequency (PMW)	B List bigram frequency	C List bigram frequency	B list word length	C list word length
	1	rat	shusk	flea	2.27	1.00	1.27	32.61	3.31	8251	7136	3	4
	1	wolf	brenk	cat	4.86	3.12	1.74	20.27	66.33	2433	7204	4	3
	1	quail	dewel	roach	3.02	1.20	1.82	1.33	2.65	4227	7183	5	5
	1	goat	plave	penguin	4.57	3.90	0.67	10.53	2.88	5749	17606	4	7
	1	cow	soble	eagle	6.47	3.82	2.65	25.51	11.49	4043	8212	3	5
	1	alligator	tenic	beaver	5.51	3.61	1.9	3.49	4.82	21597	13602	9	6
	1	donkey	badet	porcupine	5.92	3.53	2.39	5.35	0.65	8055	18884	6	9
	1	camel	gorby	duck	7.43	3.08	4.35	5.02	24.76	7872	2443	5	4
	1	rhinoceros	geter	dove	7.76	2.51	5.25	0.75	5.57	23521	3973	10	4
	1	elk	hesty	squirrel	6.33	2.61	3.72	6	5.47	2494	10121	3	9
	2	frog	smead	snail	1.98	1.20	0.78	11.82	1.76	4064	5083	4	5
	2	crab	hoxey	butterfly	1.98	1.35	0.63	6.9	5.51	5938	19071	4	9
	2	panther	hober	fox	5.63	3.98	1.65	2.57	21.61	20203	1019	7	3
	2	giraffe	bleem	tiger	7.69	5.90	1.79	1.49	18.53	7052	15108	7	5
	2	leopard	morax	mouse	5.59	1.98	3.61	5.41	19.12	11856	9286	7	5
	2	zebra	gatin	snake	6.22	3.02	3.2	2.51	22.35	4698	3510	5	5
	2	octopus	zoron	lobster	5.82	2.51	3.31	1.94	7.33	7737	18543	7	7
	2	moose	cherk	rooster	7.06	3.06	4	5.53	3.86	6202	21139	5	7
	2	bear	slend	bee	7.37	1.22	6.15	57.41	10.35	7450	2302	4	3
	2	horse	snafe	rabbit	6.65	3.06	3.59	92.88	20.94	8514	8785	5	6
Mean	1	N/A	N/A	N/A	5.41	2.84	2.58	11.09	12.79	8824.20	9636.40	5.20	5.60
	2	N/A	N/A	N/A	5.60	2.73	2.87	18.85	13.14	8371.40	10384.60	5.50	5.50
Grand Mean	**Both**	**N/A**	**N/A**	**N/A**	**1.4359**	**0.9467**	**1.26685**	**15.385**	**10.28885**	**4278.58**	**5399**	**1.555**	**1.56**

Appendix 8

List counterbalancing for Experiments 3, 4 and 5.

SET 1

CONDITION	SET	AB list	AB NW relationship	NW	AC NW relationship	AC list
Congruent	2	wolf	>	brenk	>	cat
Congruent	2	quail	>	dewel	>	roach
Congruent	2	alligator	>	tenic	>	beaver
Congruent	2	donkey	>	badet	>	porcupine
Congruent	2	elk	>	hesty	>	squirrel
Congruent	2	snail	<	smead	<	frog
Congruent	2	tiger	<	bleem	<	giraffe
Congruent	2	mouse	<	morax	<	leopard
Congruent	2	rooster	<	cherk	<	moose
Congruent	2	bee	<	slend	<	bear
Incongruent	1	rat	<	shusk	<	flea
Incongruent	1	goat	<	plave	<	penguin
Incongruent	1	cow	<	soble	<	eagle
Incongruent	1	camel	<	gorby	<	duck
Incongruent	1	rhinoceros	<	geter	<	dove
Incongruent	1	butterfly	>	hoxey	>	crab
Incongruent	1	fox	>	hober	>	panther
Incongruent	1	snake	>	gatin	>	zebra
Incongruent	1	lobster	>	zoron	>	octopus
Incongruent	1	rabbit	>	snafe	>	horse

SET 2

CONDITION	SET	AB list	AB NW relationship	NW	AC NW relationship	AC list
Congruent	2	cat	<	brenk	<	wolf
Congruent	2	roach	<	dewel	<	quail
Congruent	2	beaver	<	tenic	<	alligator
Congruent	2	porcupine	<	badet	<	donkey
Congruent	2	squirrel	<	hesty	<	elk
Congruent	2	frog	>	smead	>	snail
Congruent	2	giraffe	>	bleem	>	tiger
Congruent	2	leopard	>	morax	>	mouse
Congruent	2	moose	>	cherk	>	rooster
Congruent	2	bear	>	slend	>	bee
Incongruent	1	flea	>	shusk	>	rat
Incongruent	1	penguin	>	plave	>	goat
Incongruent	1	eagle	>	soble	>	cow
Incongruent	1	duck	>	gorby	>	camel
Incongruent	1	dove	>	geter	>	rhinoceros
Incongruent	1	crab	<	hoxey	<	butterfly
Incongruent	1	panther	<	hober	<	fox
Incongruent	1	zebra	<	gatin	<	snake
Incongruent	1	octopus	<	zoron	<	lobster
Incongruent	1	horse	<	snafe	<	rabbit

SET 3

CONDITION	SET	AB list	AB NW relationship	NW	AC NW relationship	AC list
Congruent	1	rat	>	shusk	>	flea
Congruent	1	goat	>	plave	>	penguin
Congruent	1	cow	>	soble	>	eagle
Congruent	1	camel	>	gorby	>	duck
Congruent	1	rhinoceros	>	geter	>	dove
Congruent	1	butterfly	<	hoxey	<	crab
Congruent	1	fox	<	hober	<	panther
Congruent	1	snake	<	gatin	<	zebra
Congruent	1	lobster	<	zoron	<	octopus
Congruent	1	rabbit	<	snafe	<	horse
Incongruent	2	wolf	<	brenk	<	cat
Incongruent	2	quail	<	dewel	<	roach
Incongruent	2	alligator	<	tenic	<	beaver
Incongruent	2	donkey	<	badet	<	porcupine
Incongruent	2	elk	<	hesty	<	squirrel
Incongruent	2	snail	>	smead	>	frog
Incongruent	2	tiger	>	bleem	>	giraffe
Incongruent	2	mouse	>	morax	>	leopard
Incongruent	2	rooster	>	cherk	>	moose
Incongruent	2	bee	>	slend	>	bear

SET 4

CONDITION	SET	AB list	AB NW relationship	NW	AC NW relationship	AC list
Congruent	1	flea	<	shusk	<	rat
Congruent	1	penguin	<	plave	<	goat
Congruent	1	eagle	<	soble	<	cow
Congruent	1	duck	<	gorby	<	camel
Congruent	1	dove	<	geter	<	rhinoceros
Congruent	1	crab	>	hoxey	>	butterfly
Congruent	1	panther	>	hober	>	fox
Congruent	1	zebra	>	gatin	>	snake
Congruent	1	octopus	>	zoron	>	lobster
Congruent	1	horse	>	snafe	>	rabbit
Incongruent	2	cat	>	brenk	>	wolf
Incongruent	2	roach	>	dewel	>	quail
Incongruent	2	beaver	>	tenic	>	alligator
Incongruent	2	porcupine	>	badet	>	donkey
Incongruent	2	squirrel	>	hesty	>	elk
Incongruent	2	frog	<	smead	<	snail
Incongruent	2	giraffe	<	bleem	<	tiger
Incongruent	2	leopard	<	morax	<	mouse
Incongruent	2	moose	<	cherk	<	rooster
Incongruent	2	bear	<	slend	<	bee

Appendix 9

Instructional pamphlet used in Experiments 3, 4 and 5.

Welcome to the experiment. Please read the following very carefully. Once you have read these instructions you will be asked some questions about them by the experimenter to check whether you have understood them.

The ultimate objective of this experiment is to for you to learn the size of new animals, you have not seen, in relation to animals you already know.

The new animals will be referred to by name only. During the first part of this experiment you will be asked to learn the names of these new animals and type them in. For example you might see the following:

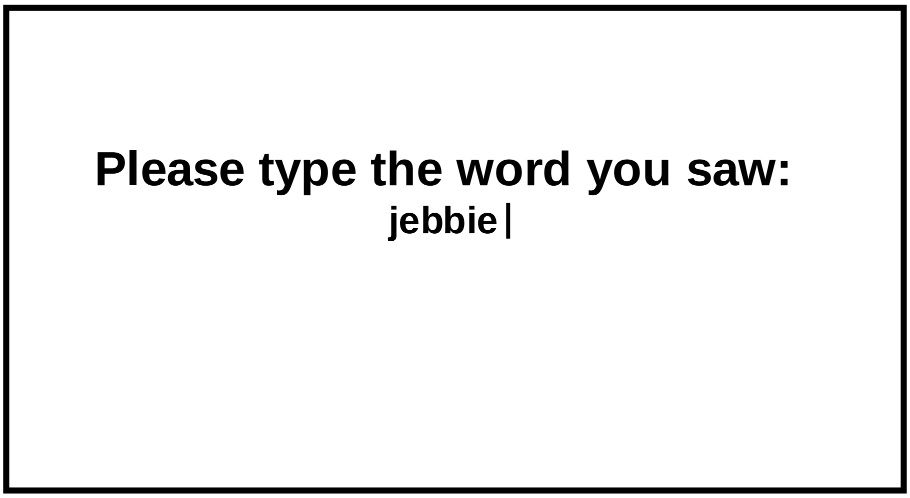
jebbie

Followed by…

Please type the word you saw:

Appendix 9 continued

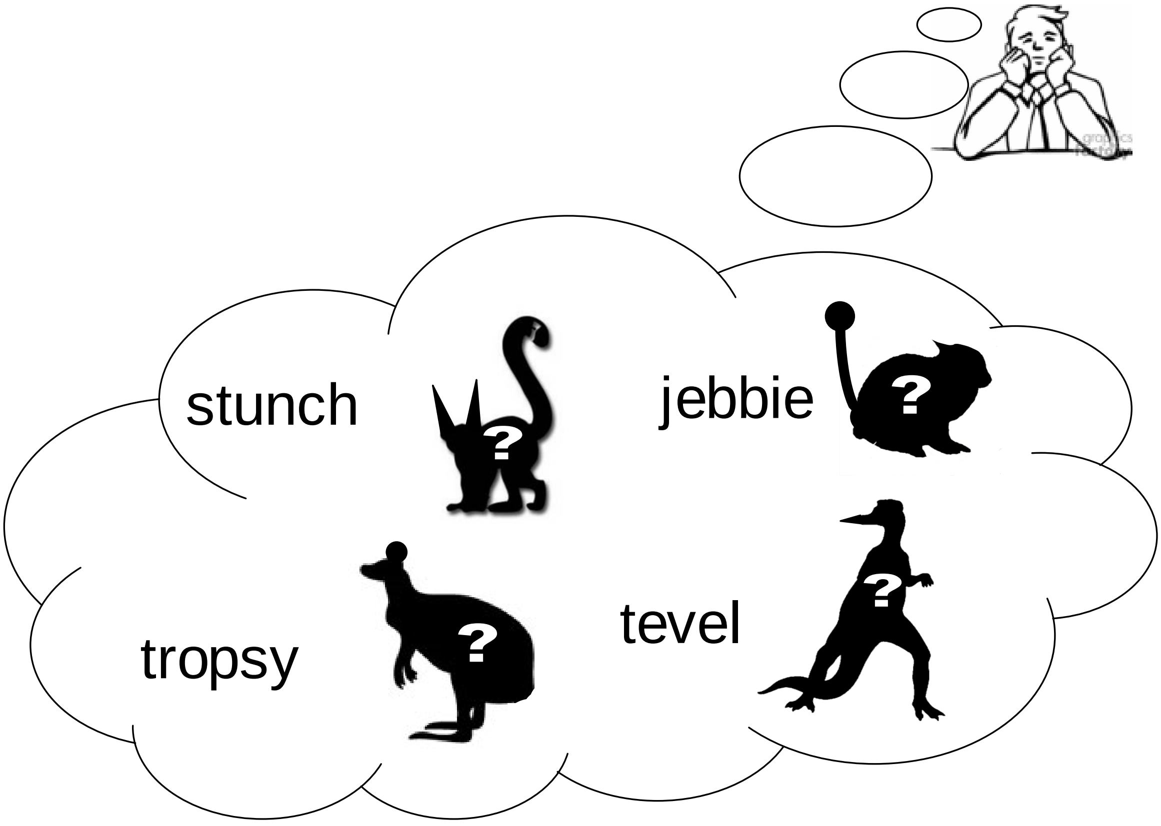
When instructed simply type the name of the word you saw as quickly and as accurately as possible followed by the return key.



Please type the word you saw:
jebbie|

then...

Despite not being able to see these new animals, or knowing what size they are, please try to imagine what they might look like. You will be asked to learn twenty new animals in this manner.

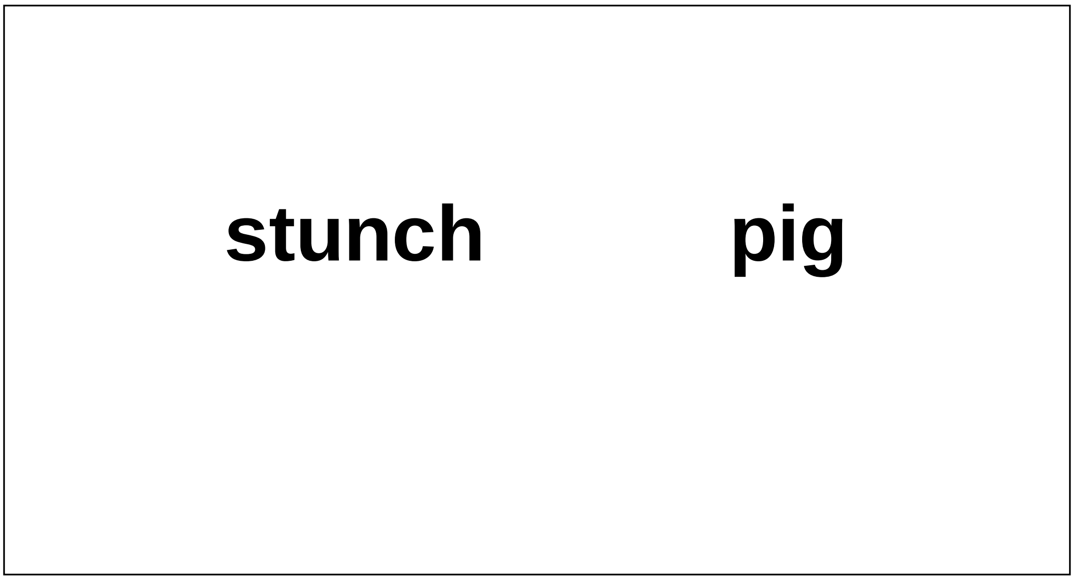


Appendix 9 continued

Once you have become familiarised with the names of the new animals you will be asked to learn what **size** they are.

Your task will be to learn the size of the new animal in relation to the animal you already know. You will see a number of animal pairs and will eventually be tested on your memory of them.

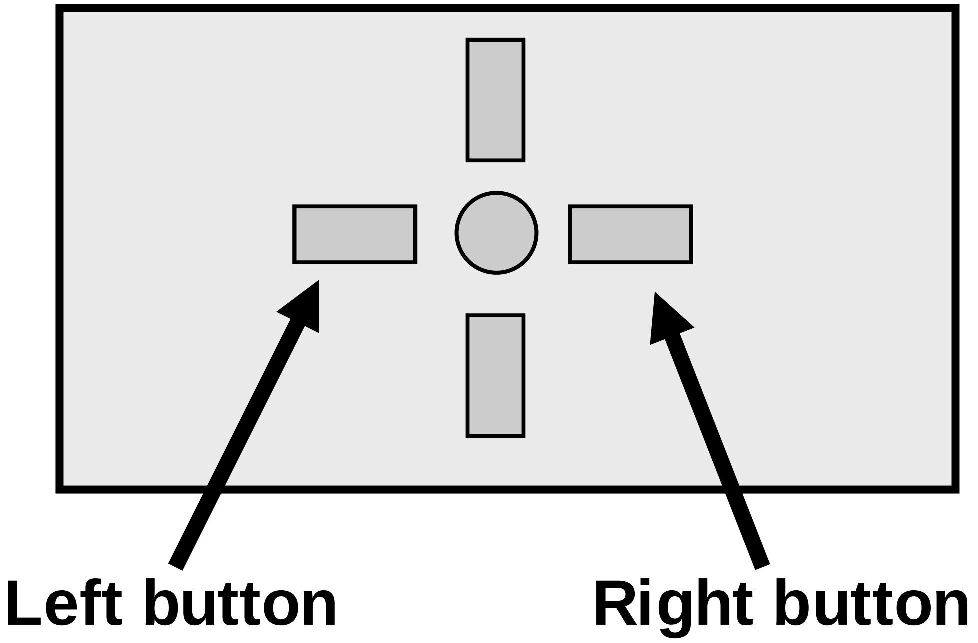
During this part of the experiment you will see a number of word pairs. One word in each pair will be an animal word you already know (eg. ‘pig’) and another will be one of the new words you were just trained on (e.g. ‘stunch’). So for example you might see the following on screen:



stunch **pig**

Appendix 9 continued

To decide which out of the two words is bigger you will use the button box provided. You will mainly be using the left and right buttons.



These buttons will correspond to the side of the screen with the animal that you think is the **bigger** of a given pair.

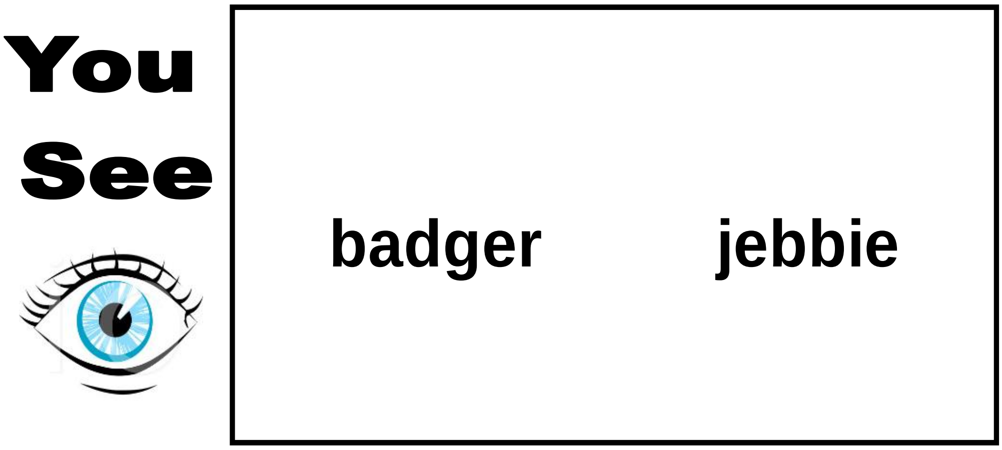
When making decisions it is important you try and visualise what size the new animal is in relation to the existing animal you already know.

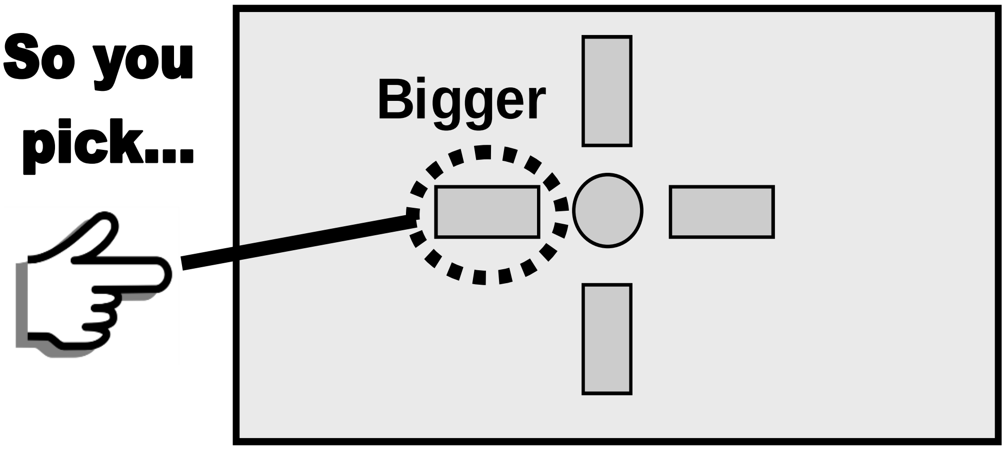
To start with this will be guess work but you will soon become a lot better with practice.

Appendix 9 continued

EXAMPLE 1:

You are presented with the following words on screen. You will be asked to visualise the items and use the button box to select the side of the screen for the animal you think is biggest in a pair. For example:

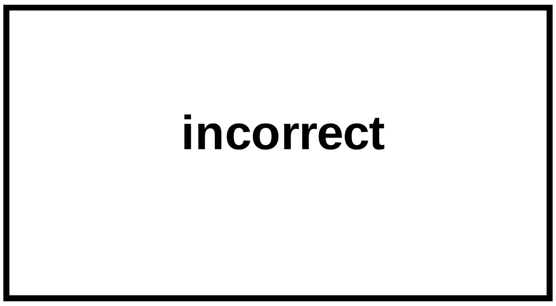




Appendix 9 continued

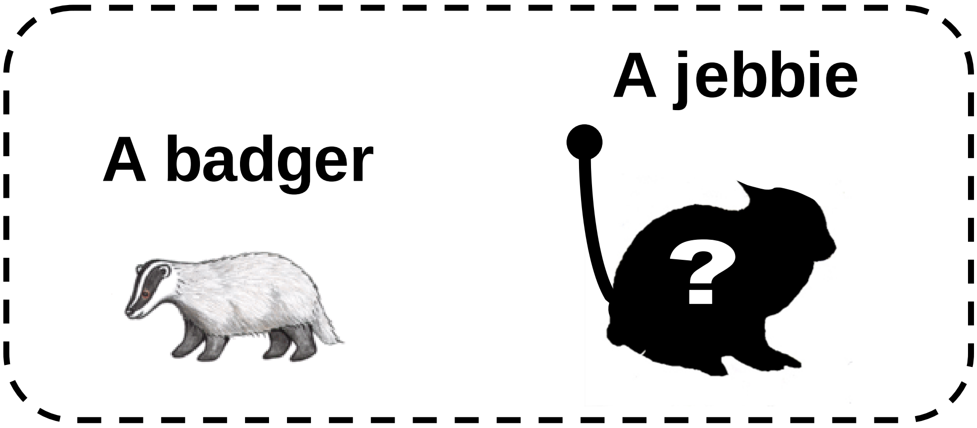
EXAMPLE 1:

You will then be shown whether this decision was correct of not. In this example you see:

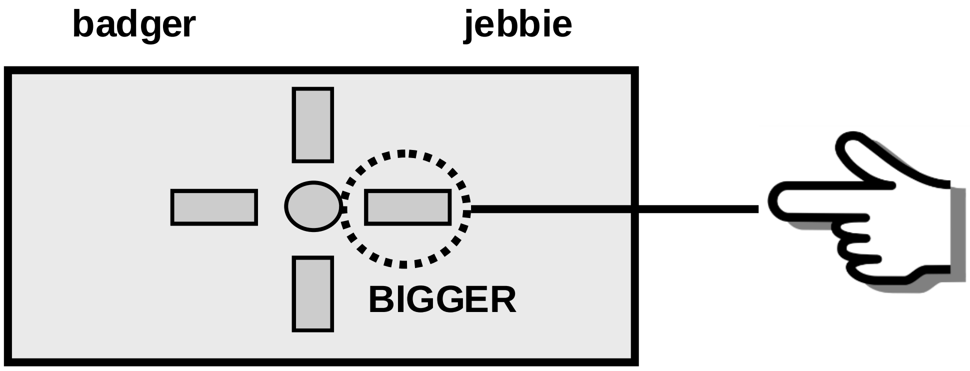


incorrect

This means your original decision was wrong, and in fact a jebbie is bigger than a badger. You should re-visualise the size of a jebbie in relation to a badger.



To get the answer right next time you see the two animals on screen you would choose the button corresponding to the side of the screen with ‘jebbie’ on it.



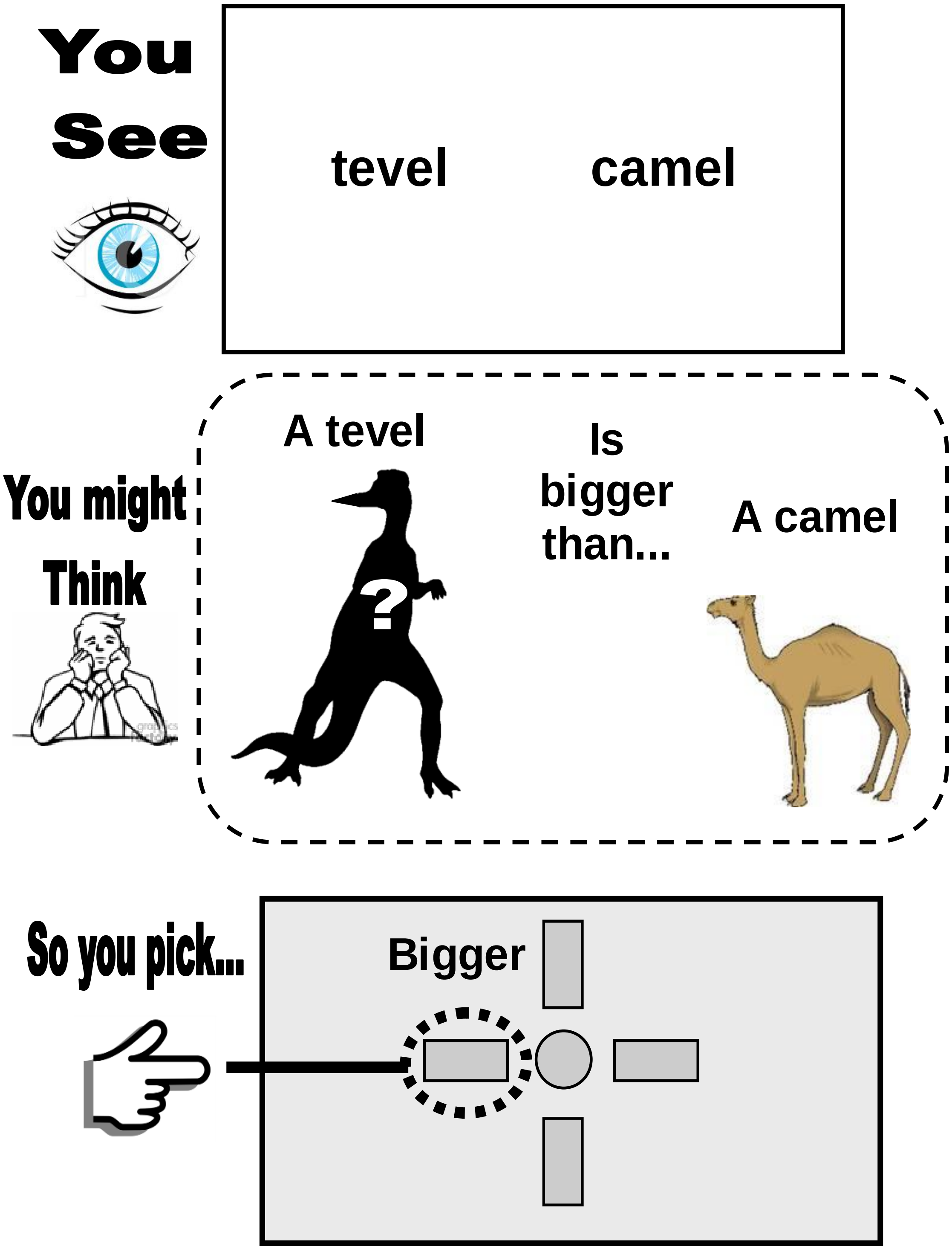
Be warned: the side of the screen that the animals will appear will alternate so don’t just press the same button each time the pair comes up!

Appendix 9 continued

EXAMPLE 2:

In this example you have seen the following animals.

Remember to try and visualise the animals when making decisions.



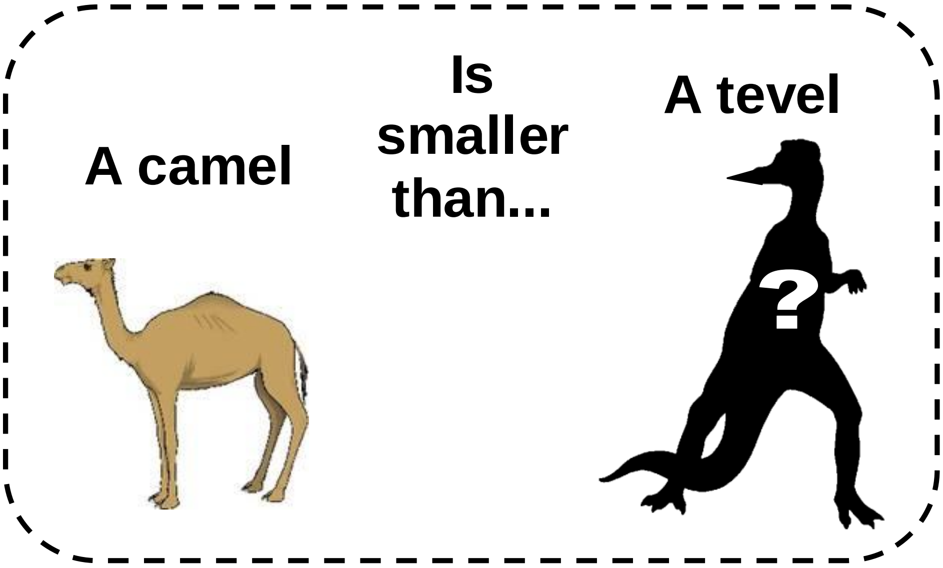
Appendix 9 continued

EXAMPLE 2:

Your decision is followed by:

correct

This means your original decision was indeed correct and you should continue to think, and visualise, a tevel as being bigger than a camel even if the words switch sides of the screen.



Do not be disheartened by getting the answer repeatedly wrong. You will soon become familiar with the bigger animal in pairs of words. The important thing is that you remain aware of the relationship between the two. You will be given feedback as you go along to help.

Please talk to the experimenter once you think you have understood these instructions.

Appendix 10

The five standardized questions participants were asked in experiments 3, 4 and 5 after reading the training pamphlet in Appendix 9.

1. What will you first be asked to learn?

2. As well as typing words when familiarising yourself with the new words what else should you do?

3. What to the left and right buttons on the button box do?

4. What should you do when you get an answer wrong?

5. In terms of which side of the screen animals will appear, I should be aware of what?

Appendix 11

The cued recall questionnaire used in Experiments 3, 4 and 5.

Over the next few pages you will see the names of the new animals you learned previously. Please try and recall the two animals that accompanied them when you were learning about their sizes. Write these animals in the spaces provided.

Please try and remember the list that these animals belonged to. For example if an animal item occurred with the new animal in the first or second list of words you learned. You can leave the space blank if you do not know the answer but feel free to guess.

You will have as long as you need to answer.

Please turn over to begin.

Appendix 11 continued

tenic

List 1:.......................................

List 2:.......................................

shusk

List 1:.......................................

List 2:.......................................

brenk

List 1:.......................................

List 2:.......................................

Appendix 11 continued

dewel

List 1:......................................

List 2:......................................

plave

List 1:......................................

List 2:......................................

soble

List 1:......................................

List 2:......................................

Appendix 11 continued

badet

List 1:.....................................

List 2:.....................................

gorby

List 1:.....................................

List 2:.....................................

geter

List 1:.....................................

List 2:.....................................

Appendix 11 continued

hesty

List 1:.......................................

List 2:.......................................

smead

List 1:.......................................

List 2:.......................................

hoxey

List 1:.......................................

List 2:.......................................

Appendix 11 continued

hober

List 1:.....................................

List 2:.....................................

bleem

List 1:.....................................

List 2:.....................................

morax

List 1:.....................................

List 2:.....................................

Appendix 11 continued

gatin

List 1:.......................................

List 2:.......................................

zoron

List 1:.......................................

List 2:.......................................

cherk

List 1:.......................................

List 2:.......................................

slend

List 1:.......................................

List 2:.......................................

snafe

List 1:.......................................

List 2:.......................................

Thanks!

Please hand this to the experimenter when you are done.

Appendix 12

Correlations between sleep stage activity and selected behavioural factors in Experiment 3.

The symbol '†' denotes p-values that do not survive a Bonferroni correction for multiple comparisons. Corrections in this case involve an inflation of the alpha value by the total number of behavioural factors. Please refer to section 4.3.5 for more details.

Cued recall (AB-AC)		TST	REM %	Light sleep %	SWS %
Congruent correct List %	*r*	.454	**.640**	-.258	-.526
	p	*.161*	***.034†***	*.443*	*.096*
Incongruent correct List %	*r*	-.039	.151	-.117	-.056
	p	*.909*	*.659*	*.732*	*.871*
Congruent incorrect List %	*r*	.411	.431	-.562	.106
	p	*.209*	*.186*	*.072*	*.757*
Incongruent incorrect List %	*r*	.305	.131	-483	.412
	p	*.362*	*.701*	*.132*	*.207*
2AFC list judgements (AB-AC)					
Congruent accuracy %	*r*	**.665**	.377	-.019	-.478
	p	***.025†***	*.253*	*.957*	*1.37*
Incongruent accuracy %	*r*	.014	-.013	.145	-.156
	p	*.967*	*.969*	*.671*	*.629*
Congruent RT (msec)	*r*	.252	-.096	-.149	.302
	p	*.455*	*.779*	*.663*	*.367*
Incongruent RT (msec)	*r*	.075	.174	.259	-.537
	p	*.826*	*.609*	*.441*	*.089*
2AFC size generalisations (Narrow - Wide)					
Congruent generalisation (%)	*r*	-.154	-.159	.533	-.431
	p	*.650*	*.640*	*.091*	*.186*
Incongruent generalisation (%)	*r*	-.446	-.394	.095	.418
	p	*.169*	*.230*	*.782*	*.201*
Congruent RT (msec)	*r*	-.154	-.159	.533	-.431
	p	*.650*	*.640*	*.091*	*.186*
Incongruent RT (msec)	*r*	-.446	-.394	.095	.418
	p	*.169*	*.230*	*.782*	*.201*

Note: Significant correlations in bold. TST =total sleep time, REM% = percentage of time spent in rapid eye movement sleep, Light sleep % = percentage of time spent in stages 1 and 2 sleep combined, SWS %= percentage of time spent in slow wave sleep. † = p-values that do not survive a Bonferroni correction for multiple comparisons. Behavioural measures represent composite values (see text).

Appendix 13

Correlations between sleep stage activity and selected behavioural factors in Experiment 4.

Corrections in this case involve an inflation of the alpha value by the total number of behavioural factors. Please refer to section 4.3.5 for more details.

Cued recall (AB-AC)		TST	REM %	Light sleep %	SWS %
Congruent correct List %	*r*	-.123	.218	-.193	.024
	p	*.753*	*.573*	*.618*	*.951*
Incongruent correct List %	*r*	-.131	-.210	.313	-.149
	p	*.737*	*.588*	*.412*	*.702*
Congruent incorrect List %	*r*	.245	-.094	.026	.077
	p	*.526*	*.810*	*.947*	*.844*
Incongruent incorrect List %	*r*	-.434	-.216	.134	.075
	p	*.244*	*.577*	*.732*	*.849*
2AFC list judgements (AB-AC)					
Congruent accuracy %	*r*	.219	-.453	.277	.208
	p	*.572*	*.221*	*.471*	*.592*
Incongruent accuracy %	*r*	.294	.297	-.467	.260
	p	*.443*	*.437*	*.205*	*.498*
Congruent RT (msec)	*r*	-.435	-.211	.541	-.478
	p	*.242*	*.585*	*.133*	*.193*
Incongruent RT (msec)	*r*	.397	-.470	.163	.392
	p	*.290*	*.202*	*.675*	*.297*
2AFC size generalisations (Narrow- Wide)					
Congruent generalisation (%)	*r*	.156	.038	.005	-.101
	p	*.688*	*.922*	*.991*	*.796*
Incongruent generalisation (%)	*r*	-.343	.256	-.156	-.126
	p	*.366*	*.507*	*.688*	*.746*
Congruent RT (msec)	*r*	-.233	-.050	.151	-.131
	p	*.547*	*.898*	*.699*	*.737*
Incongruent RT (msec)	*r*	.478	.264	-.615	.512
	p	*.193*	*.492*	*.078*	*.159*

Note: Significant correlations in bold. TST =total sleep time, REM% = percentage of time spent in rapid eye movement sleep, Light sleep % = percentage of time spent in stages 1 and 2 sleep combined, SWS %= percentage of time spent in slow wave sleep. p-values are reported without correction for multiple comparisons. Behavioural measures represent composite values .

Appendix 14

Correlations between sleep stage % activity and selected behavioural factors in Experiment 5.

Cued recall (AB-AC)		TST (mins)	REM %	Stage 2 %	SWS %
Congruent correct List %	*r*	.034	.151	-.110	-.312
	p	*.904*	*.591*	*.696*	*.257*
Incongruent correct List %	*r*	.171	.117	.014	-.006
	p	*.543*	*.527*	*.960*	*.984*
2AFC list judgements					
A-B Congruent accuracy %	*r*	-.159	-.013	.071	-.122
	p	*.573*	*.963*	*.803*	*.666*
A-C Congruent accuracy %	*r*	-.380	-.225	-.057	-.111
	p	*.162*	*.421*	*.841*	*.694*
A-B Incongruent accuracy %	*r*	-.059	-.084	-.048	-.177
	p	*.833*	*.765*	*.864*	*.529*
A-C Incongruent accuracy %	*r*	-.572	-.506	.163	.112
	p	*.026*	*.054*	*.561*	*.692*
A-B Congruent RT (msec)	*r*	-.171	-.226	.385	-.146
	p	*.543*	*.418*	*.157*	*.604*
A-C Congruent RT (msec)	*r*	-.115	-.100	.095	.165
	p	*.684*	*.722*	*.737*	*.556*
A-B Incongruent RT (msec)	*r*	-.123	-.103	.180	-.168
	p	*.661*	*.714*	*.520*	*.551*
A-C Incongruent RT (msec)	*r*	-.309	-.375	.448	-.280
	p	*.262*	*.169*	*.094*	*.312*
2AFC size generalisations (Narrow- Wide)					
Congruent generalisation (%)	*r*	-.358	-.387	.378	-.388
	p	*.191*	*.154*	*.164*	*.153*
Incongruent generalisation (%)	*r*	-.473	-.466	.358	-.403
	p	*.075*	*.080*	*.191*	*.137*
Congruent RT (msec)	*r*	.128	.210	-.039	.184
	p	*.672*	*.542*	*.889*	*.511*
Incongruent RT (msec)	*r*	.126	-.001	.087	-.224
	p	*.655*	*.996*	*.889*	*.423*

Note: Significant correlations in bold. TST =total sleep time, REM% = percentage of time spent in rapid eye movement sleep, SWS %= percentage of time spent in slow wave sleep. p-values are reported without correction for multiple comparisons.

Appendix 15

Correlations between sleep stage time activity and selected behavioural factors in Experiment 5.

Cued recall (AB-AC)		REM time (mins)	Stage 1 time (mins)	SWS time (mins)
Congruent correct List %	*r*	.034	-.187	-.384
	p	*.158*	*.504*	*.158*
Incongruent correct List %	*r*	.171	.022	.000
	p	*.534*	*.937*	*.999*
2AFC list judgements				
A-B Congruent accuracy %	*r*	-.159	-.060	-.209
	p	*.573*	*.833*	*.454*
A-C Congruent accuracy %	*r*	-.380	-.205	-.234
	p	*.162*	*.463*	*.402*
A-B Incongruent accuracy %	*r*	-.059	-.001	-.149
	p	*.833*	*.997*	*.596*
A-C Incongruent accuracy %	*r*	**-.572**	.062	.018
	p	***.026†***	*.826*	*.948*
A-B Congruent RT (msec)	*r*	-.171	.379	-.102
	p	*.543*	*.164*	*.717*
A-C Congruent RT (msec)	*r*	-.115	.089	.171
	p	*.684*	*.753*	*.543*
A-B Incongruent RT (msec)	*r*	-.123	.142	-.169
	p	*.661*	*.614*	*.547*
A-C Incongruent RT (msec)	*r*	-.309	.447	-.231
	p	*.262*	*.095*	*.407*
2AFC size generalisations (Narrow- Wide)				
Congruent generalisation (%)	*r*	-.358	.276	-.398
	p	*.191*	*.319*	*.142*
Incongruent generalisation (%)	*r*	-.473	.242	-.457
	p	*.075*	*.384*	*.087*
Congruent RT (msec)	*r*	.128	-.097	.136
	p	*.649*	*.731*	*.629*
Incongruent RT (msec)	*r*	.126	.159	-.148
	p	*.655*	*.571*	*.599*

Note: Significant correlations in bold. TST =total sleep time, REM% = percentage of time spent in rapid eye movement sleep, SWS %= percentage of time spent in slow wave sleep. p-values are reported without correction for multiple comparisons. The correlation marked with the (*†*) symbol did not survive correction for multiple comparisons.

Appendix 16

Spindle density across the NREM sleep period correlated with selected behavioural factors in Experiment 5.

Cued recall (AB-AC)		All spindles (11-15 Hz)	Fast spindles (13-15 Hz)	Slow spindles (11-15 Hz)
Congruent correct List %	*r*	-.080	-.204	.022
	p	*.776*	*.465*	*.938*
Incongruent correct List %	*r*	-.395	-.460	-.215
	p	*.145*	*.084*	*.441*
2AFC list judgements				
A-B Congruent accuracy %	*r*	-.033	-.250	.305
	p	*.908*	*.369*	*.269*
A-C Congruent accuracy %	*r*	-.159	-.464	-.026
	p	*.571*	*.082*	*.926*
A-B Incongruent accuracy %	*r*	.175	-.022	.172
	p	*.533*	*.938*	*.540*
A-C Incongruent accuracy %	*r*	.215	-.012	.340
	p	*.442*	*.965*	*.216*
A-B Congruent RT (msec)	*r*	.116	-.001	.275
	p	*.681*	*.997*	*.322*
A-C Congruent RT (msec)	*r*	-.060	-.156	.147
	p	*.830*	*.580*	*.600*
A-B Incongruent RT (msec)	*r*	-.090	-.309	.052
	p	*.751*	*.262*	*.853*
A-C Incongruent RT (msec)	*r*	.026	-.131	.182
	p	*.928*	*.641*	*.517*
2AFC size generalisations (Narrow- Wide)				
Congruent generalisation (%)	*r*	.307	.298	.101
	p	*.265*	*.281*	*.720*
Incongruent generalisation (%)	*r*	.282	.120	.167
	p	*.308*	*.671*	*.553*
Congruent RT (msec)	*r*	-.066	-.123	.147
	p	*.816*	*.662*	*.601*
Incongruent RT (msec)	*r*	-.070	.029	-.214
	p	*.805*	*.920*	*.444*

Note: Behavioural factors are correlated with spindle density (spindles per minute) with three measures of spindle density aggregated across the entire Non-rapid eye movement (NREM) sleep period. SWS = Slow wave sleep. P-values are reported unadjusted for multiple comparison (see text).

Appendix 17

Correlations between global spindle density (11-15 Hz) with the 14 behavioural factors extracted from the behavioural measures featured in Experiment 5.

			Stage 2 sleep		Slow Wave Sleep	
Cued recall (AB-AC)			Central	Frontal	Central	Frontal
	Congruent correct List %	*r*	-.037	-.034	-.304	-.082
		p	*.897*	*.906*	*.270*	*.770*
	Incongruent correct List %	*r*	-.286	-.438	-.382	-.082
		p	*.302*	*.103*	*.160*	*.770*
2AFC list judgements						
	A-B Congruent accuracy %	*r*	-0.218	-0.065	-0.065	0.129
		p	*0.44*	*0.82*	*0.82*	*0.65*
	A-C Congruent accuracy %	*r*	0.003	-0.189	-0.168	-0.196
		p	*0.99*	*0.50*	*0.55*	*0.48*
	A-B Incongruent accuracy %	*r*	0.115	0.015	0.255	0.259
		p	*0.68*	*0.96*	*0.36*	*0.35*
	A-C Incongruent accuracy %	*r*	0.061	0.059	0.421	0.357
		p	*0.83*	*0.83*	*0.12*	*0.19*
	A-B Congruent RT (m sec)	*r*	0.055	0.134	0.060	0.176
		p	*0.85*	*0.63*	*0.83*	*0.53*
	A-C Congruent RT (m sec)	*r*	-0.175	-0.029	-0.086	0.064
		p	*0.53*	*0.92*	*0.76*	*0.82*
	A-B Incongruent RT (m sec)	*r*	-0.009	-0.138	-0.197	-0.028
		p	*0.98*	*0.62*	*0.48*	*0.92*
	A-C Incongruent RT (m sec)	*r*	-0.005	0.060	-0.055	0.100
		p	*0.99*	*0.83*	*0.85*	*0.72*
2AFC size generalisations (Narrow- Wide)						
	Congruent generalisation (%)	*r*	.301	.235	.347	.235
		p	*.276*	*.399*	*.205*	*.399*
	Incongruent generalisation (%)	*r*	.343	.183	.371	.196
		p	*.210*	*.531*	*.174-*	*.485*
	Congruent RT (m sec)	*r*	-.168	-.157	-.132	.153
		p	*.549*	*.576*	*.670*	*.586*
	Incongruent RT (m sec)	*r*	.060	-.128	-.130	-.085
		p	*.831*	*.650*	*.645*	*.763*

Note: SWS = Slow wave sleep. Behavioural factors are correlated with spindle density (spindles per minute) in the 11-15 Hz range. p-values are reported unadjusted for multiple comparisons.

Appendix 18

Correlations between fast spindle density (13-15 Hz) with the 14 behavioural factors extracted from the behavioural measures featured in Experiment 5.

		Stage 2 sleep		Slow Wave Sleep	
Cued recall (AB-AC)		Central	Frontal	Central	Frontal
Congruent correct List %	*r*	-.361	-.151	-.270	-.031
	p	*.186*	*.591*	*.331*	*.914*
Incongruent correct List %	*r*	-.349	-.491	-.384	-.258
	p	*.202*	*.063*	*.158*	*.354*
2AFC list judgements					
A-B Congruent accuracy %	*r*	**-.549**	-.400	-.071	.115
	p	***.034†***	*.140*	*.802*	*.683*
A-C Congruent accuracy %	*r*	-.157	-.260	-.514	**-.548**
	p	*.576*	*.350*	*.050*	***.034†***
A-B Incongruent accuracy %	*r*	-.029	-.238	.196	.061
	p	*.918*	*.393*	*.485*	*.830*
A-C Incongruent accuracy %	*r*	-.066	-.154	.179	.070
	p	*.816*	*.583*	*.523*	*.804*
A-B Congruent RT (msec)	*r*	-.112	-.153	.147	.156
	p	*.690*	*.585*	*.601*	*.578*
A-C Congruent RT (msec)	*r*	-.344	-.291	.036	.104
	p	*.209*	*.293*	*.898*	*.712*
A-B Incongruent RT (msec)	*r*	-.228	-.386	-.213	-.157
	p	*.413*	*.155*	*.445*	*.577*
A-C Incongruent RT (msec)	*r*	-.234	-.275	.068	.072
	p	*.401*	*.320*	*.809*	*.798*
2AFC size generalisations (Narrow- Wide)					
Congruent generalisation (%)	*r*	.499	.295	.252	-.021
	p	*.058*	*.286*	*.365*	*.942*
Incongruent generalisation (%)	*r*	.366	.146	.119	-.160
	p	*.180*	*.604*	*.672*	*.568*
Congruent RT (msec)	*r*	-.410	-.328	-.042	.283
	p	*.129*	*.233*	*.883*	*.306*
Incongruent RT (msec)	*r*	.190	.016	-.067	-.025
	p	*.497*	*.954*	*.811*	*.931*

Note: SWS = Slow wave sleep. Behavioural factors are correlated with spindle density (spindles per minute) in the 13-15 Hz range. p-values are reported unadjusted for multiple comparisons. The symbol '*†*' indicates correlations that did not survive correction for multiple comparisons.

Appendix 19

Correlations between slow spindle density (11-13 Hz) with the 14 behavioural factors extracted from the behavioural measures featured in Experiment 5.

		Stage 2 sleep		Slow Wave Sleep	
Cued recall (AB-AC)		Central	Frontal	Central	Frontal
Congruent correct List %	*r*	.245	.134	-.283	-.186
	p	*.378*	*.634*	*.306*	*.507*
Incongruent correct List %	*r*	-.114	-.271	-.276	-.160
	p	*.686*	*.328*	*.319*	*.569*
2AFC list judgements					
A-B Congruent accuracy %	*r*	.289	.310	.199	.250
	p	*.296*	*.260*	*.478*	*.368*
A-C Congruent accuracy %	*r*	.030	-.142	.092	-.010
	p	*.915*	*.613*	*.744*	*.971*
A-B Incongruent accuracy %	*r*	.030	.090	.073	.331
	p	*.917*	*.751*	*.796*	*.228*
A-C Incongruent accuracy %	*r*	.254	.230	**.539**	.387
	p	*.361*	*.410*	***.038†***	*.154*
A-B Congruent RT (msec)	*r*	.203	.254	.309	.315
	p	*.467*	*.362*	*.263*	*.253*
A-C Congruent RT (msec)	*r*	.061	.147	.173	.208
	p	*.828*	*.600*	*.539*	*.456*
A-B Incongruent RT (msec)	*r*	.029	-.039	.016	.158
	p	*.917*	*.891*	*.953*	*.573*
A-C Incongruent RT (msec)	*r*	.113	.216	.114	.244
	p	*.687*	*.439*	*.685*	*.381*
2AFC size generalisations (Narrow- Wide)					
Congruent generalisation (%)	*r*	-.189	.069	.149	.246
	p	*.499*	*.808*	*.596*	*.378*
Incongruent generalisation (%)	*r*	.065	.125	.239	.182
	p	*.817*	*.657*	*.391*	*.517*
Congruent RT (msec)	*r*	.301	.113	.018	.095
	p	*.275*	*.689*	*.948*	*.735*
Incongruent RT (msec)	*r*	-.153	-.272	-.280	-.126
	p	*.587*	*.327*	*.313*	*.654*

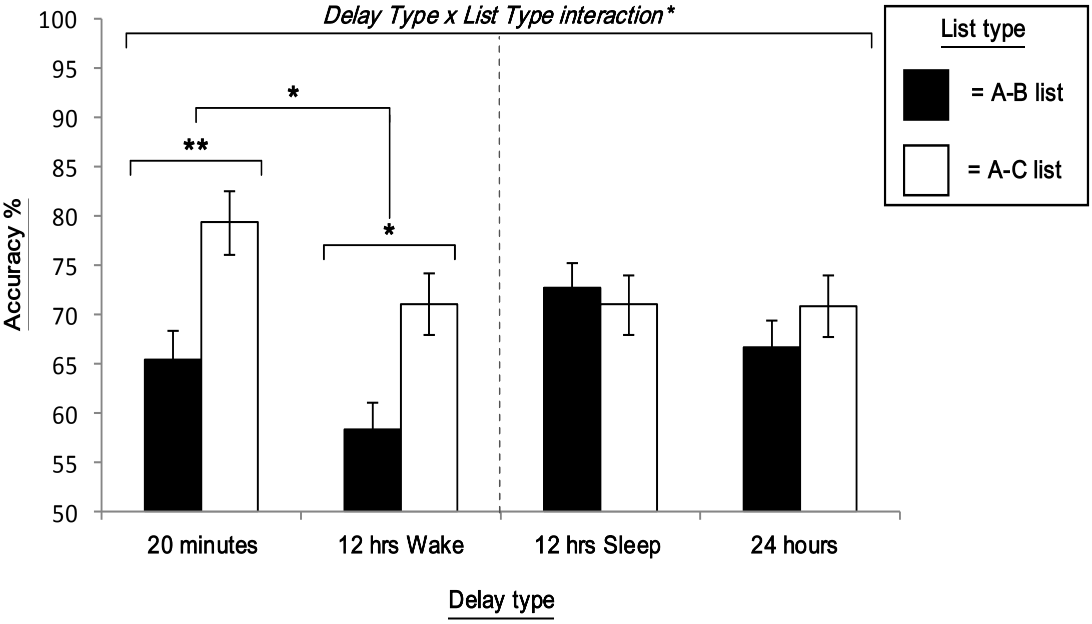
Note: SWS = Slow wave sleep. Behavioural factors are correlated with spindle density (spindles per minute) in the 11-13 Hz range. p-values are reported unadjusted for multiple comparisons. The symbol '*†*' indicates correlations that did not survive correction for multiple comparisons.

Appendix 20

Supplementary analyses of the aggregated Congruent list accuracy measures across the four Experimental time points featured in Experiments 3 and 5.

MAIN ANALYSIS:

A 4x2 mixed-effects ANOVA featuring the between-participant variable of Delay type (Immediate, Wake, Sleep, or Delayed) and the with-participants variable of List-type (A-B list vs. A-C list). The dependent variable was Congruent list accuracy (%).



RESULTS:

- A main effect of List type was evident, *F*(1,62)=12.74, *p*=.001, driven by lower A-B list accuracy (*M*=66%, *SE*=1.44) relative to the A-C list (*M*=73%, *SE*=1.56)
- A marginal main effect of Delay type was evident, *F*(3, 62)=2.69, *p*=.055.
- An interaction emerged between Delay type and List, *F*(3,62)=3.3, *p*=.026 (see below).

FOLLOW UP ANALYSES:

- Paired samples t-tests indicate RI effects between the A-B and A-C lists in both the Immediate, *t*(14)=-4.36, *p*=.001, and Wake conditions, *t*(16)=-2.55, *p*=.022. These RI effects disappeared after a consolidation opportunity, in the Sleep, *t*<1, and Delayed, *t*(15)=-1.1, *p*=.287, conditions. As such a consolidation opportunity appears to remove RI effects.
- There was also a drop in average accuracy performance between the Immediate (*M*=72%, *SE*=2.66) and Wake (*M*=65%, *SE*=1.73) conditions, *t*(30)=2.48, *p*=.019. However, no such drop in accuracy was seen between the Sleep and Delayed conditions, *t*(32)=1.00, *p*=.324, indicating that memories were protected from waking interference after a consolidation-opportunity.

Appendix 21

Power analyses of PSG data and the resulting correlation coefficients contained in this thesis.

Statistical power:

- Type I error (False positive) = α was set at .05 by default for all analyses.
- Type II error (False negative) = β should be ≥ .80 to detect a real effect (Cohen, 1988).
- The following power calculations below are post hoc based on the uncorrected coefficients reported in Chapters 3, 4, and 5.
- In experiments with no marginally significant correlations (experiments 4 and 5) the equivalent correlations are reported for the last previous marginally significant set of result (as seen for Experiment 3).
- The software used to calculate the power (β-1) was G*Power.

Experiment	**Number of PSG records**	**Correlation**	**R value**	**P value**	**Power (1-β error prob)**
Experiment 2	n=14	SWS (min) vs. 2 Unit TI (%)	0.585	.028	0.64
		REM (min) vs. Real word inference (msec)	-.552	.041	0.57
		REM (min) vs. 2 Unit TI (msec)	-0.564	0.36	0.59
		TST (min) vs. RW Size diff (%)	-0.547	.043	0.56
		SWS (msec) vs. NW Size diff (msec)	-0.586	.028	0.64
		SWS spindles (%) vs. Premises (%)	0.603	.022	0.67
		SWS spindles (%) vs. NW inferences (%)	0.580	0.30	0.63
		S2 Spindles (%) vs. 2 Unit TI (%)	0.599	.024	0.66
Experiment 3	n=11	REM% vs. Congruent correct list (%)	.640	.034	0.61
		TST (min) 2AFC Congruent list (%)	.665	.025	0.66
Experiment 4	n=9	REM% vs. Congruent correct list (%)	.218	.573	0.09
		TST (min) 2AFC Congruent list (%)	.219	-.453	0.09
Experiment 5	n=15	REM% vs. Congruent correct list (%)	.151	591	0.08
		TST (min) 2AFC A-B Congruent list (%)	-.159	.573	0.09

Conclusions:

- In general the levels of power for the sleep recordings did not reach the minimum levels of suggested power (0.8, Cohen, 1988).
- This was expressed most prominently for Experiments 3-4.
- Future experiments would be best served by a power analysis prior to the experiment to determine the number of suitable participants required.

References

Abel, M., & Bäuml, K.-H. T. (2013). Sleep can reduce proactive interference. *Memory (Hove, England)*. doi:10.1080/09658211.2013.785570

Abra, J. C. (1969). List-1 unlearning and recovery as a function of the point of interpolated learning. *Journal of Verbal Learning and Verbal Behavior*, *8*(4), 494–500. doi:10.1016/S0022-5371(69)80093-9

Acuna, B., Sanes, J., & Donoghue, J. (2002). Cognitive mechanisms of transitive inference. *Experimental Brain Research*, *146*(1), 128. doi:10.1007/s00221-002-1226-2

Agnew, H. W., Webb, W. B., & Williams, R. L. (1966). The First Night Effect: An EEG Study of Sleep. *Psychophysiology*, *2*(3), 263–266. doi:10.1111/j.1469-8986.1966.tb02650.x

Agranoff, B. W., Davis, R. E., & Brink, J. J. (1966). Chemical studies on memory fixation in goldfish. *Brain Research*, *1*(3), 303–309.

Aldrich, M. S. (1999). *Sleep Medicine* (1st ed.). Oxford University Press, USA.

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

Anderer, P., Klösch, G., Gruber, G., Trenker, E., Pascual-Marqui, R. ., Zeitlhofer, J., … Saletu, B. (2001). Low-resolution brain electromagnetic tomography revealed simultaneously active frontal and parietal sleep spindle sources in the human cortex. *Neuroscience*, *103*(3), 581–592. doi:10.1016/S0306-4522(01)00028-8

Anderson. (1983). Retrieval of information from long-term memory. *Science*, *220*(4592), 25–30. doi:10.1126/science.6828877

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*(4), 415–445. doi:10.1016/j.jml.2003.08.006

Andrillon, T., Nir, Y., Staba, R. J., Ferrarelli, F., Cirelli, C., Tononi, G., & Fried, I. (2011). Sleep Spindles in Humans: Insights from Intracranial EEG and Unit Recordings. *The Journal of Neuroscience*, *31*(49), 17821–17834. doi:10.1523/JNEUROSCI.2604-11.2011

Aserinsky, E., & Kleitman, N. (1953). Regularly Occurring Periods of Eye Motility, and Concomitant Phenomena, During Sleep. *Science*, *118*, 273–274. doi:10.1176/appi.neuropsych.15.4.454

Atienza, M., Cantero, J. L., & Stickgold, R. (2004). Posttraining Sleep Enhances Automaticity in Perceptual Discrimination. *Journal of Cognitive Neuroscience*, *16*(1), 53–64. doi:10.1162/089892904322755557

Baddeley, A. D., & Dale, H. C. A. (1966). The effect of semantic similarity on retroactive interference in long- and short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *5*(5), 417–420. doi:10.1016/S0022-5371(66)80054-3

Barnes, J. M., & Underwood, B. J. (1959). 'Fate' of first-list associations in transfer theory. *Journal of Experimental Psychology*, *58*(2), 97–105. doi:10.1037/h0047507

Barr, W. B., Goldberg, E., Wasserstein, J., & Novelly, R. A. (1990). Retrograde amnesia following unilateral temporal lobectomy. *Neuropsychologia*, *28*(3), 243–255.

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. doi:10.1037/h0033625

Bell, M. C., Kawadri, N., Simone, P. M., & Wiseheart, M. (2013). Long-term memory, sleep, and the spacing effect. *Memory*, *0*(0), 1–8. doi:10.1080/09658211.2013.778294

Bendor, D., & Wilson, M. A. (2012). Biasing the content of hippocampal replay during sleep. *Nature Neuroscience*, *15*(10), 1439–1444. doi:10.1038/nn.3203

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. doi:10.2307/2346101

Benson, K., & Feinberg, I. (1977). The Beneficial Effect of Sleep in an Extended Jenkins and Dallenbach Paradigm. *Psychophysiology*, *14*(4), 375–384. doi:10.1111/j.1469-8986.1977.tb02967.x

Blake, M. J. (1967). Time of day effects on performance in a range of tasks. *Psychonomic Science*, *9*(6), 349–350.

Bontempi, B., Laurent-Demir, C., Destrade, C., & Jaffard, R. (1999). Time-dependent reorganization of brain circuitry underlying long-term memory storage. *Nature*, *400*(6745), 671–675. doi:10.1038/23270

Borbély, A. A., & Achermann, P. (1999). Sleep homeostasis and models of sleep regulation. *Journal of Biological Rhythms*, *14*(6), 557–568.

Born, J., & Gais, S. (2000). REM Sleep Deprivation: The Wrong Paradigm Leading to Wrong Conclusions. *Behavioral and Brain Sciences*, *23*(06), 912–913. doi:null

Born, J., Rasch, B., & Gais, S. (2006). Sleep to Remember. *The Neuroscientist*, *12*(5), 410–424. doi:10.1177/1073858406292647

Born, J., & Wilhelm, I. (2012). System consolidation of memory during sleep. *Psychological Research*, *76*(2), 192–203. doi:10.1007/s00426-011-0335-6

Bower, G. H., Thompson-Schill, S., & Tulving, E. (1994). Reducing retroactive interference: An interference analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(1), 51–66. doi:10.1037/0278-7393.20.1.51

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-Trace Theory and False Memory. *Current Directions in Psychological Science*, *11*(5), 164–169. doi:10.1111/1467-8721.00192

Brashers-Krug, T., Shadmehr, R., & Bizzi, E. (1996). Consolidation in human motor memory. *Nature*, *382*(6588), 252–255. doi:10.1038/382252a0

Braun, A. R., Balkin, T. J., Wesenten, N. J., Carson, R. E., Varga, M., Baldwin, P., … Herscovitch, P. (1997). Regional cerebral blood flow throughout the sleep-wake cycle. An H2(15)O PET study. *Brain*, *120*(7), 1173–1197. doi:10.1093/brain/120.7.1173

Brown, R. M., & Robertson, E. M. (2007). Off-Line Processing: Reciprocal Interactions between Declarative and Procedural Memories. *The Journal of Neuroscience*, *27*(39), 10468 –10475. doi:10.1523/JNEUROSCI.2799-07.2007

Bryant, P. E., & Trabasso, T. (1971). Transitive Inferences and Memory in Young Children. *Nature*, *232*(5311), 456–458. doi:10.1038/232456a0

Buxton, C. E. (1943). The status of research in reminiscence. *Psychological Bulletin*, *40*(5), 313–340. doi:10.1037/h0053614

Buzsáki, G. (1989). Two-stage model of memory trace formation: A role for 'noisy' brain states. *Neuroscience*, *31*(3), 551–570. doi:10.1016/0306-4522(89)90423-5

Buzsáki, G. (1996). The Hippocampo-Neocortical Dialogue. *Cerebral Cortex*, *6*(2), 81 –92. doi:10.1093/cercor/6.2.81

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. doi:10.1073/pnas.0900271106

Cairney, S. (2012). *Exploring the mechanisms and specificity of emotional memory consolidation over sleep*.

Ceraso, J., & Henderson, A. (1965). Unavailability and associative loss in in RI and PI. *Journal of Experimental Psychology*, *70*(3), 300–303. doi:10.1037/h0022247

Christianson, S., & Christianson, S.-A. (1992). *The Handbook of Emotion and Memory: Research and Theory*. Routledge.

Cipolli, C. (1995). Sleep, Dreams and memory: an overview. *Journal of Sleep Research*, *4*(1), 2–9. doi:10.1111/j.1365-2869.1995.tb00143.x

Cipolotti, L., Shallice, T., Chan, D., Fox, N., Scahill, R., Harrison, G., … Rudge, P. (2001). Long-term retrograde amnesia… the crucial role of the hippocampus. *Neuropsychologia*, *39*(2), 151–172. doi:10.1016/S0028-3932(00)00103-2

Cirelli, C., & Tononi, G. (2000). Differential Expression of Plasticity-Related Genes in Waking and Sleep and Their Regulation by the Noradrenergic System. *The Journal of Neuroscience*, *20*(24), 9187–9194.

Clemens, Z., Fabó, D., & Halász, P. (2005). Overnight verbal memory retention correlates with the number of sleep spindles. *Neuroscience*, *132*(2), 529–535. doi:16/j.neuroscience.2005.01.011

Clemens, Z., Mölle, M., Erőss, L., Barsi, P., Halász, P., & Born, J. (2007). Temporal coupling of parahippocampal ripples, sleep spindles and slow oscillations in humans. *Brain*, *130*(11), 2868–2878. doi:10.1093/brain/awm146

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Psychology Press.

Conte, F., & Ficca, G. (2012). Caveats on psychological models of sleep and memory: A compass in an overgrown scenario. *Sleep Medicine Reviews*. doi:10.1016/j.smrv.2012.04.001

Coutanche, M. N., Gianessi, C. A., Chanales, A. J. H., Willison, K. W., & Thompson-Schill, S. L. (2013). The Role of Sleep in Forming a Memory Representation of a Two-dimensional Space. *Hippocampus*, (In press), 1–17.

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

Danker-Hopfe, H., Anderer, P., Zeitlhofer, J., Boeck, M., Dorn, H., Gruber, G., … Dorffner, G. (2009). Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard. *Journal of Sleep Research*, *18*(1), 74–84. doi:10.1111/j.1365-2869.2008.00700.x

Darsaud, A., Dehon, H., Lahl, O., Sterpenich, V., Boly, M., Dang-Vu, T., … Collette, F. (2010). Does Sleep Promote False Memories? *Journal of Cognitive Neuroscience*, *23*(1), 26–40. doi:10.1162/jocn.2010.21448

Dean, G. M., Dewhurst, S. A., Morris, P. E., & Whittaker, A. (2005). Selective Interference With the Use of Visual Images in the Symbolic Distance Paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 1043–1068. doi:10.1037/0278-7393.31.5.1043

Delaney, H. D., & Maxwell, S. E. (1981). On Using Analysis Of Covariance In Repeated Measures Designs. *Multivariate Behavioral Research*, *16*(1), 105–123. doi:10.1207/s15327906mbr1601_6

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. doi:10.1037/0033-295X.93.3.283

Diekelmann, S., & Born, J. (2010a). The memory function of sleep. *Nat Rev Neurosci*, *11*(2), 114–126. doi:10.1038/nrn2762

Diekelmann, S., & Born, J. (2010b, March 1). Slow-wave sleep takes the leading role in memory reorganization. Nature Publishing Group.

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

Diekelmann, S., Landolt, H.-P., Lahl, O., Born, J., & Wagner, U. (2008). Sleep Loss Produces False Memories. *PLoS ONE*, *3*(10), e3512. doi:10.1371/journal.pone.0003512

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

Diekelmann, S., Wilhelm, I., Wagner, U., & Born, J. (2013). Sleep to implement an intention. *Sleep*, *36*(1), 149–153. doi:10.5665/sleep.2322

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. doi:10.1037/0096-3445.136.2.169

Drosopoulos, S., Windau, E., Wagner, U., & Born, J. (2007). Sleep Enforces the Temporal Order in Memory. *PLoS ONE*, *2*(4), e376. doi:10.1371/journal.pone.0000376

Dudai, Y. (1996). Consolidation: fragility on the road to the engram. *Neuron*, *17*(3), 367–370.

Dudai, Y. (2004). The Neurobiology of Consolidations, Or, How Stable is the Engram? *Annual Review of Psychology*, *55*(1), 51–86. doi:10.1146/annurev.psych.55.090902.142050

Dudai, Y. (2012). The Restless Engram: Consolidations Never End. *Annual Review of Neuroscience*, *35*(1), 227–247. doi:10.1146/annurev-neuro-062111-150500

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

Dumay, N., Gaskell, M. G., & Feng, X. (2004). A day in the life of a spoken word. Conference or workshop item. Retrieved 5 January 2011, from http://kar.kent.ac.uk/14939/

Duncan, C. P. (1945). The effect of electroshock convulsions on the maze habit in the white rat. *Journal of Experimental Psychology*, *35*(4), 267–278. doi:10.1037/h0059587

Duncan, C. P. (1949). The retroactive effect of electroshock on learning. *Journal of Comparative and Physiological Psychology*, *42*(1), 32–44. doi:10.1037/h0058173

Durrant, S. J., Cairney, S. A., & Lewis, P. A. (2012). Overnight Consolidation Aids the Transfer of Statistical Knowledge from the Medial Temporal Lobe to the Striatum. *Cerebral Cortex (New York, N.Y.: 1991)*. doi:10.1093/cercor/bhs244

Durrant, S. J., Taylor, C., Cairney, S., & Lewis, P. A. (2011). Sleep-dependent consolidation of statistical learning. *Neuropsychologia*, *49*(5), 1322–1331. doi:10.1016/j.neuropsychologia.2011.02.015

Dusek, J. A., & Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(13), 7109 –7114.

Duyck, W., Desmet, T., Verbeke, L. P. C., & Brysbaert, M. (2004). WordGen: A tool for word selection and nonword generation in Dutch, English, German, and French. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 488–499. doi:VL - 36

Ebbinghaus, H. (1885). *Memory: A Contribution to Experimental Psychology*. (H. A. Ruger & C. E. Bussenius, Trans.). Teachers College, Columbia University.

Eichenbaum, H. (2004). Hippocampus: Cognitive Processes and Neural Representations that Underlie Declarative Memory. *Neuron*, *44*(1), 109–120. doi:10.1016/j.neuron.2004.08.028

Eichenbaum, H. (2010). Declarative Memory. In *The Corsini Encyclopedia of Psychology*. John Wiley & Sons, Inc. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/9780470479216.corpsy0253/abstract

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

Ekstrand, B. R., Sullivan, M. J., Parker, D. F., & West, J. N. (1971). Spontaneous recovery and sleep. *Journal of Experimental Psychology*, *88*(1), 142–144. doi:10.1037/h0030642

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. doi:10.1073/pnas.0700094104

Ellenbogen, J. M., Hulbert, J. C., Jiang, Y., & Stickgold, R. (2009). The Sleeping Brain's Influence on Verbal Memory: Boosting Resistance to Interference. *PLoS ONE*, *4*(1), e4117. doi:10.1371/journal.pone.0004117

Ellenbogen, J. M., Hulbert, J. C., Stickgold, R., Dinges, D. F., & Thompson-Schill, S. L. (2006). Interfering with Theories of Sleep and Memory: Sleep, Declarative Memory, and Associative Interference. *Current Biology*, *16*(13), 1290–1294. doi:10.1016/j.cub.2006.05.024

Ellenbogen, J. M., Payne, J. D., & Stickgold, R. (2006). The role of sleep in declarative memory consolidation: passive, permissive, active or none? *Current Opinion in Neurobiology*, *16*(6), 716–722. doi:10.1016/j.conb.2006.10.006

Empson, J. a. C., & Clarke, P. R. F. (1970). Rapid Eye Movements and Remembering. *, Published Online: 18 July 1970; | doi:10.1038/227287a0*, *227*(5255), 287–288. doi:10.1038/227287a0

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, *120*(4), 339–357. doi:10.1037/0096-3445.120.4.339

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

Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature*, *425*(6958), 614–616. doi:10.1038/nature01951

Ferrarelli, F., Huber, R., Peterson, M. J., Massimini, M., Murphy, M., Riedner, B. A., … Tononi, G. (2007). Reduced Sleep Spindle Activity in Schizophrenia Patients. *Am J Psychiatry*, *164*(3), 483–492. doi:<p>10.1176/appi.ajp.164.3.483</p>

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. doi:10.1016/S0166-4328(00)00177-7

Field, A. (2009). *Discovering Statistics Using SPSS* (Third Edition.). Sage Publications Ltd.

Fischer, S., Drosopoulos, S., Tsen, J., & Born, J. (2006). Implicit Learning–Explicit Knowing: A Role for Sleep in Memory System Interaction. *Journal of Cognitive Neuroscience*, *18*(3), 311–319. doi:i: 10.1162/jocn.2006.18.3.311</p>

Fischer, S., Hallschmid, M., Elsner, A. L., & Born, J. (2002). Sleep forms memory for finger skills. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(18), 11987 –11991. doi:10.1073/pnas.182178199

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. doi:10.1016/j.bbr.2007.02.037

Foulkes, D., Hollifield, M., Sullivan, B., Bradley, L., & Terry, R. (1990). REM Dreaming and Cognitive Skills at Ages 5-8: A Cross-sectional Study. *International Journal of Behavioral Development*, *13*(4), 447 –465. doi:10.1177/016502549001300404

Fowler, M. J., Sullivan, M. J., & Ekstrand, B. R. (1973). Sleep and Memory. *Science*, *179*(4070), 302 –304. doi:10.1126/science.179.4070.302

Frank, M. J., O'Reilly, R. C., & Curran, T. (2006). When Memory Fails, Intuition Reigns. *Psychological Science*, *17*(8), 700 –707. doi:10.1111/j.1467-9280.2006.01769.x

Frank, M. J., Rudy, J. W., Levy, W. B., & O'Reilly, R. C. (2005). When logic fails: implicit transitive inference in humans. *Memory and Cognition*, *4*(33), 742–750.

Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nat Rev Neurosci*, *6*(2), 119–130. doi:10.1038/nrn1607

Frankland, P. W., & Bontempi, B. (2006). Fast track to the medial prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(3), 509 –510. doi:10.1073/pnas.0510133103

Frankland, P. W., Bontempi, B., Talton, L. E., Kaczmarek, L., & Silva, A. J. (2004). The Involvement of the Anterior Cingulate Cortex in Remote Contextual Fear Memory. *Science*, *304*(5672), 881–883. doi:10.1126/science.1094804

Frankland, P. W., O'Brien, C., Ohno, M., Kirkwood, A., & Silva, A. J. (2001). α-CaMKII-dependent plasticity in the cortex is required for permanent memory. *Nature*, *411*(6835), 309–313. doi:10.1038/35077089

French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, *3*(4), 128–135. doi:10.1016/S1364-6613(99)01294-2

Gais, S., Albouy, G., Boly, M., Dang-Vu, T. T., Darsaud, A., Desseilles, M., … Peigneux, P. (2007). Sleep transforms the cerebral trace of declarative memories. *Proceedings of the National Academy of Sciences*, *104*(47), 18778 –18783. doi:10.1073/pnas.0705454104

Gais, S., Lucas, B., & Born, J. (2006). Sleep after learning aids memory recall. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *13*(3), 259–262. doi:10.1101/lm.132106

Gais, S., Molle, M., Helms, K., & Born, J. (2002). Learning-Dependent Increases in Sleep Spindle Density. *J. Neurosci.*, *22*(15), 6830–6834. doi:20026697

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *38*(7), 833–848. doi:10.3758/MC.38.7.833

Gaskell, G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105–132. doi:10.1016/S0010-0277(03)00070-2

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. doi:10.1016/0166-4328(95)00012-I

Göder, R., & Born, J. (2012). Can sleep heal memory? *Sleep Medicine Reviews*, (0). doi:10.1016/j.smrv.2012.09.001

Goerke, M., Cohrs, S., Rodenbeck, A., Grittner, U., Sommer, W., & Kunz, D. (2012). Declarative memory consolidation during the first night in a sleep lab: The role of REM sleep and cortisol. *Psychoneuroendocrinology*. doi:10.1016/j.psyneuen.2012.10.019

Gómez, R. L., Bootzin, R. R., & Nadel, L. (2006). Naps Promote Abstraction in Language-Learning Infants. *Psychological Science*, *17*(8), 670–674. doi:10.1111/j.1467-9280.2006.01764.x

Graham, K. S., Becker, J. T., & Hodges, J. R. (1997). On the relationship between knowledge and memory for pictures: Evidence from the study of patients with semantic dementia and Alzheimer's disease. *Journal of the International Neuropsychological Society*, *3*(06), 534–544.

Greene, A. J. (2007). Implicit transitive inference and the human hippocampus: does intravenous midazolam function as a reversible hippocampal lesion? *Behavioral and Brain Functions*, *3*(1), 51. doi:10.1186/1744-9081-3-51

Greene, A. J., Gross, W. L., Elsinger, C. L., & Rao, S. M. (2006). An fMRI Analysis of the Human Hippocampus: Inference, Context, and Task Awareness. *Journal of Cognitive Neuroscience*, *18*(7), 1156–1173. doi:10.1162/jocn.2006.18.7.1156

Greene, A. J., Spellman, B. A., Dusek, J. A., Eichenbaum, H. B., & Levy, W. B. (2001). Relational learning with and without awareness: transitive inference using nonverbal stimuli in humans. *Memory & Cognition*, *29*(6), 893–902.

Griessenberger, H., Heib, D. P. J., Kunz, A. B., Hoedlmoser, K., & Schabus, M. (2012). Assessment of a wireless headband for automatic sleep scoring. *Sleep and Breathing*, 1–6. doi:10.1007/s11325-012-0757-4

Groeger, J. A., Lo, J. C. Y., Burns, C. G., & Dijk, D.-J. (2011). Effects of sleep inertia after daytime naps vary with executive load and time of day. *Behavioral Neuroscience*, *125*(2), 252–260. doi:10.1037/a0022692

Grosenick, L., Clement, T. S., & Fernald, R. D. (2007). Fish can infer social rank by observation alone. *Nature*, *445*(7126), 429–432. doi:10.1038/nature05511

Grosvenor, A., & Lack, L. C. (1984). The effect of sleep before or after learning on memory. *Sleep*, *7*(2), 155–167.

Hartley, D. (1834). *Observations on man, his frame, his duty, and his expectations*. Printed for T. Tegg and sons.

Hauptmann, B., Reinhart, E., Brandt, S. A., & Karni, A. (2005). The predictive value of the leveling off of within session performance for procedural memory consolidation. *Cognitive Brain Research*, *24*(2), 181–189. doi:10.1016/j.cogbrainres.2005.01.012

Healy, A. F., Kosslyn, S. M., & Shiffrin, R. M. (1992). *From Learning Theory to Connectionist Theory: Essays in Honor of William K. Estes*. Routledge.

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory* (New edition.). Psychology Press.

Heckers, S., Zalesak, M., Weiss, A. P., Ditman, T., & Titone, D. (2004). Hippocampal activation during transitive inference in humans. *Hippocampus*, *14*(2), 153–162. doi:10.1002/hipo.10189

Henderson, L. M., Weighall, A. R., Brown, H., & Gaskell, G. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science*, *15*(5), 674–687. doi:10.1111/j.1467-7687.2012.01172.x

Himanen, S. L., & Hasan, J. (2000). Limitations of Rechtschaffen and Kales. *Sleep Medicine Reviews*, *4*(2), 149–167. doi:10.1053/smrv.1999.0086

Hobson, J. A., & Pace-Schott, E. F. (2002). The cognitive neuroscience of sleep: neuronal systems, consciousness and learning. *Nature Reviews Neuroscience*, *3*(9), 679–693. doi:10.1038/nrn915

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

Huber, R., Ghilardi, M. F., Massimini, M., Ferrarelli, F., Riedner, B. A., Peterson, M. J., & Tononi, G. (2006). Arm immobilization causes cortical plastic changes and locally decreases sleep slow wave activity. *Nature Neuroscience*, *9*(9), 1169–1176. doi:10.1038/nn1758

Huber, R., Ghilardi, M. F., Massimini, M., & Tononi, G. (2004). Local sleep and learning. *Nature*, *430*(6995), 78–81. doi:10.1038/nature02663

Inostroza, M., & Born, J. (2013). Sleep for Preserving and Transforming Episodic Memory. *Annual Review of Neuroscience*. doi:10.1146/annurev-neuro-062012-170429

Jasper, H. H. (1958). The ten twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, *10*, 371–375.

Jenkins, J. G., & Dallenbach, K. M. (1924). Obliviscence during Sleep and Waking. *The American Journal of Psychology*, *35*(4), 605–612.

Jones, M. W., & Wilson, M. A. (2005). Theta Rhythms Coordinate Hippocampal–Prefrontal Interactions in a Spatial Memory Task. *PLoS Biol*, *3*(12), e402. doi:10.1371/journal.pbio.0030402

Karni, A., Tanne, D., Rubenstein, B., Askenasy, J., & Sagi, D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. *Science*, *265*(5172), 679 –682. doi:10.1126/science.8036518

Kim, J. J., & Fanselow, M. S. (1992). Modality-specific retrograde amnesia of fear. *Science*, *256*(5057), 675–677. doi:10.1126/science.1585183

King, C., Henze, D. A., Leinekugel, X., & Buzsáki, G. (1999). Hebbian modification of a hippocampal population pattern in the rat. *The Journal of Physiology*, *521*(1), 159–167. doi:10.1111/j.1469-7793.1999.00159.x

Knott, G. W., Quairiaux, C., Genoud, C., & Welker, E. (2002). Formation of Dendritic Spines with GABAergic Synapses Induced by Whisker Stimulation in Adult Mice. *Neuron*, *34*(2), 265–273. doi:10.1016/S0896-6273(02)00663-3

Konkel, A., & Cohen, N. J. (2009). Relational Memory and the Hippocampus: Representations and Methods. *Frontiers in Neuroscience*, *3*(2), 166–174. doi:10.3389/neuro.01.023.2009

Koppenaal, R. J. (1963). Time changes in the strengths of A–B, A–C lists; spontaneous recovery? *Journal of Verbal Learning and Verbal Behavior*, *2*(4), 310–319. doi:10.1016/S0022-5371(63)80099-7

Korman, M., Doyon, J., Doljansky, J., Carrier, J., Dagan, Y., & Karni, A. (2007). Daytime sleep condenses the time course of motor memory consolidation. *Nat Neurosci*, *10*(9), 1206–1213. doi:10.1038/nn1959

Kroes, M. C. W., & Fernández, G. (2012). Dynamic neural systems enable adaptive, flexible memories. *Neuroscience and Biobehavioral Reviews*, *36*(7), 1646–1666. doi:10.1016/j.neubiorev.2012.02.014

Kuhl, B. A., Shah, A. T., DuBrow, S., & Wagner, A. D. (2010). Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. *Nature Neuroscience*, *13*(4), 501–506. doi:10.1038/nn.2498

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. doi:10.1037/a0028681

Kumaran, D., Summerfield, J. J., Hassabis, D., & Maguire, E. A. (2009). Tracking the Emergence of Conceptual Knowledge during Human Decision Making. *Neuron*, *63*(6), 889–901. doi:10.1016/j.neuron.2009.07.030

Kuriyama, K., Stickgold, R., & Walker, M. P. (2004). Sleep-dependent learning and motor-skill complexity. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *11*(6), 705–713. doi:10.1101/lm.76304

Lahl, O., Wispel, C., Willigens, B., & Pietrowsky, R. (2008). An ultra short episode of sleep is sufficient to promote declarative memory performance. *Journal of Sleep Research*, *17*(1), 3–10. doi:10.1111/j.1365-2869.2008.00622.x

Lau, H., Alger, S. E., & Fishbein, W. (2011). Relational Memory: A Daytime Nap Facilitates the Abstraction of General Concepts. *PLoS ONE*, *6*(11), e27139. doi:10.1371/journal.pone.0027139

Lau, H., Tucker, M. A., & Fishbein, W. (2010). Daytime napping: Effects on human direct associative and relational memory. *Neurobiology of Learning and Memory*, *In Press, Corrected Proof*. doi:10.1016/j.nlm.2010.02.003

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, *55*(4), 306–353. doi:10.1016/j.cogpsych.2007.01.001

Lechner, H. A., Squire, L. R., & Byrne, J. H. (1999). 100 Years of Consolidation—Remembering Müller and Pilzecker. *Learning & Memory*, *6*(2), 77–87. doi:10.1101/lm.6.2.77

Lewin, I., & Glaubman, H. (1975). The Effect of REM Deprivation: Is It Detrimental, Beneficial, or Neutral? *Psychophysiology*, *12*(3), 349–353. doi:10.1111/j.1469-8986.1975.tb01303.x

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

Libben, M., & Titone, D. (2008). The role of awareness and working memory in human transitive inference. *Behavioural Processes*, *77*(1), 43–54. doi:10.1016/j.beproc.2007.06.006

Lindsay, S., & Gaskell, M. G. (2012). Lexical Integration of Novel Words Without Sleep. *Journal of Experimental Psychology. Learning, Memory, and Cognition*. doi:10.1037/a0029243

MacKinnon, D. F., & Squire, L. R. (1989). Autobiographical memory and amnesia. *Psychobiology*, *17*(3), 247–256.

Maguire, E. A., Henson, R. N. A., Mummery, C. J., & Frith, C. D. (2001). Activity in prefrontal cortex, not hippocampus, varies parametrically with the increasing remoteness of memories. *Neuroreport*, *12*(3), 441–444. doi:10.1097/00001756-200103050-00004

Mandai, O., Guerrien, A., Sockeel, P., Dujardin, K., & Leconte, P. (1989). REM sleep modifications following a Morse code learning session in humans. *Physiology & Behavior*, *46*(4), 639–642.

Maquet, P. (2000). Functional neuroimaging of normal human sleep by positron emission tomography. *Journal of Sleep Research*, *9*(3), 207–231.

Maquet, P., Péters, J.-M., Aerts, J., Delfiore, G., Degueldre, C., Luxen, A., & Franck, G. (1996). Functional neuroanatomy of human rapid-eye-movement sleep and dreaming. *Nature*, *383*, 163–166. doi:10.1038/383163a0

Marr, D. (1970). A Theory for Cerebral Neocortex. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *176*(1043), 161–234. doi:10.1098/rspb.1970.0040

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

Marshall, L., Helgadottir, H., Molle, M., & Born, J. (2006). Boosting slow oscillations during sleep potentiates memory. *Nature*, *444*(7119), 610–613. doi:10.1038/nature05278

Marshall, L., Kirov, R., Brade, J., Mölle, M., & Born, J. (2011). Transcranial Electrical Currents to Probe EEG Brain Rhythms and Memory Consolidation during Sleep in Humans. *PLoS ONE*, *6*(2), e16905. doi:10.1371/journal.pone.0016905

Martin, A. (1999). Automatic activation of the medial temporal lobe during encoding: Lateralized influences of meaning and novelty. *Hippocampus*, *9*(1), 62–70. doi:10.1002/(SICI)1098-1063(1999)9:1<62::AID-HIPO7>3.0.CO;2-K

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. doi:10.1046/j.1365-2869.1999.00154.x

McClelland, J. L., & Goddard, N. H. (1996). Considerations arising from a complementary learning systems perspective on hippocampus and neocortex. *Hippocampus*, *6*(6), 654–665. doi:10.1002/(SICI)1098-1063(1996)6:6<654::AID-HIPO8>3.0.CO;2-G

McClelland, J. L., McNaughton, B. L., & O’Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. doi:10.1037/0033-295X.102.3.419

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews. Neuroscience*, *4*(4), 310–322. doi:10.1038/nrn1076

McGaugh, J. L. (2000). Memory--a century of consolidation. *Science (New York, N.Y.)*, *287*(5451), 248–251.

McGeoch, J. A., & Irion, A. L. (1952). *The psychology of human learning. (2nd ed.)* (Vol. xxii). Oxford, England: Longmans, Green & Co.

McGonigle, B., & Chalmers, M. (1977). Are monkeys logical? *Nature*, *267*, 694–696.

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

Mednick, S. C., McDevitt, E. A., Walsh, J. K., Wamsley, E., Paulus, M., Kanady, J. C., & Drummond, S. P. A. (2013). The Critical Role of Sleep Spindles in Hippocampal-Dependent Memory: A Pharmacology Study. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(10), 4494–4504. doi:10.1523/JNEUROSCI.3127-12.2013

Mednick, S. C., Nakayama, K., Cantero, J. L., Atienza, M., Levin, A. A., Pathak, N., & Stickgold, R. (2002). The restorative effect of naps on perceptual deterioration. *Nature Neuroscience*, *5*(7), 677–681. doi:10.1038/nn864

Mednick, S. C., Nakayama, K., & Stickgold, R. (2003). Sleep-dependent learning: a nap is as good as a night. *Nat Neurosci*, *6*(7), 697–698. doi:10.1038/nn1078

Meeter, M., & Murre, J. M. J. (2004). Consolidation of long-term memory: evidence and alternatives. *Psychological Bulletin*, *130*(6), 843–857. doi:10.1037/0033-2909.130.6.843

Melton, A. W., & Irwin, J. M. (1940). The Influence of Degree of Interpolated Learning on Retroactive Inhibition and the Overt Transfer of Specific Responses. *The American Journal of Psychology*, *53*(2), 173–203. doi:10.2307/1417415

Milner, B. (1972). Disorders of learning and memory after temporal lobe lesions in man. *Clinical Neurosurgery*, *19*, 421–446.

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. doi:10.5665/SLEEP.1290

Mölle, M., Eschenko, O., Gais, S., Sara, S. J., & Born, J. (2009). The influence of learning on sleep slow oscillations and associated spindles and ripples in humans and rats. *European Journal of Neuroscience*, *29*(5), 1071–1081. doi:10.1111/j.1460-9568.2009.06654.x

Mölle, M., Marshall, L., Gais, S., & Born, J. (2002). Grouping of Spindle Activity during Slow Oscillations in Human Non-Rapid Eye Movement Sleep. *The Journal of Neuroscience*, *22*(24), 10941–10947.

Mölle, M., Marshall, L., Gais, S., & Born, J. (2004). Learning increases human electroencephalographic coherence during subsequent slow sleep oscillations. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(38), 13963 –13968. doi:10.1073/pnas.0402820101

Monk, T. H., & Leng, V. C. (1982). Time of day effects in simple repetitive tasks: Some possible mechanisms. *Acta Psychologica*, *51*(3), 207–221. doi:10.1016/0001-6918(82)90035-X

Montagna, P., & Lugaresi, E. (2002). Agrypnia Excitata: a generalized overactivity syndrome and a useful concept in the neurophysiopathology of sleep. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *113*(4), 552–560.

Moscovitch, M., & Nadel, L. (1998). Consolidation and the hippocampal complex revisited: in defense of the multiple-trace model. *Current Opinion in Neurobiology*, *8*(2), 297–300.

Moscovitch, M., Nadel, L., Winocur, G., Gilboa, A., & Rosenbaum, R. S. (2006). The cognitive neuroscience of remote episodic, semantic and spatial memory. *Current Opinion in Neurobiology*, *16*(2), 179–190. doi:10.1016/j.conb.2006.03.013

Moses, S. N., Villate, C., & Ryan, J. D. (2006). An investigation of learning strategy supporting transitive inference performance in humans compared to other species. *Neuropsychologia*, *44*(8), 1370–1387. doi:10.1016/j.neuropsychologia.2006.01.004

Moyer, R. S. (1973). Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception & Psychophysics.*, *13*(2), 180–184.

Muller, G. E., & Pilzecker, A. (1900). Experimentelle Beitra¨ge zur Lehre vom Gedachtnis. *Psychol. Erganzungsband*, *1*, 1–300.

Murphy, G. L. (2004). *The big book of concepts*. MIT Press.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, *7*(2), 217–227. doi:10.1016/S0959-4388(97)80010-4

Nadel, L., Winocur, G., Ryan, L., & Moscovitch, M. (2007). Systems consolidation and hippocampus: two views. *Debates in Neuroscience*, *1*(2), 55–66. doi:10.1007/s11559-007-9003-9

Nader, K., Schafe, G. E., & Doux, J. E. L. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, *406*(6797), 722–726. doi:10.1038/35021052

Nishida, M., & Walker, M. P. (2007). Daytime naps, motor memory consolidation and regionally specific sleep spindles. *PloS One*, *2*(4), e341. doi:10.1371/journal.pone.0000341

Norman, K. A., Newman, E. L., & Perotte, A. J. (2005). Methods for reducing interference in the Complementary Learning Systems model: Oscillating inhibition and autonomous memory rehearsal. *Neural Networks*, *18*(9), 1212–1228. doi:16/j.neunet.2005.08.010

O’Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. *Trends in Cognitive Sciences*, *6*(12), 505–510. doi:10.1016/S1364-6613(02)02005-3

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. doi:10.1037/0033-295X.108.2.311

Ohayon, M. M., Carskadon, M. A., Guilleminault, C., & Vitiello, M. V. (2004). Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. *Sleep*, *27*(7), 1255–1273.

Pace-Schott, E. F., & Hobson, J. A. (2002). The neurobiology of sleep: genetics, cellular physiology and subcortical networks. *Nature Reviews. Neuroscience*, *3*(8), 591–605. doi:10.1038/nrn895

Pavio, A. (1975). Perceptual comparisons through the mind’s eye. *Memory and Cognition*, *3*(6), 635–674.

Pavlides, C., & Winson, J. (1989). Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. *The Journal of Neuroscience*, *9*(8), 2907 –2918.

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. doi:10.1016/j.nlm.2009.03.007

Payne, J. D., Tucker, M. A., Ellenbogen, J. M., Wamsley, E. J., Walker, M. P., Schacter, D. L., & Stickgold, R. (2012). Memory for Semantically Related and Unrelated Declarative Information: The Benefit of Sleep, the Cost of Wake. *PLoS ONE*, *7*(3), e33079. doi:10.1371/journal.pone.0033079

Peters, K. R., Smith, V., & Smith, C. (2006). Sleep spindles and motor adaptation learning: the role of initial skill level. *Sleep*, *29*.

Piaget, J. (1928). *Judgment And Reasoning In The Child.* London: Routledge, Chapman & Hall.

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

Plihal, W., & Born, J. (1999). Effects of early and late nocturnal sleep on priming and spatial memory. *Psychophysiology*, *36*(5), 571–582. doi:10.1111/1469-8986.3650571

Postman, L., & Stark, K. (1969). Role of response availability in transfer and interference. *Journal of Experimental Psychology*, *79*(1, Pt.1), 168–177. doi:10.1037/h0026932

Postman, L., Stark, K., & Henschel, D. (1969). Conditions of recovery after unlearning. *Journal of Experimental Psychology*, *82*(1, Pt.2), 1–24. doi:10.1037/h0028036

Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition*, *1*(1), 19–40. doi:10.3758/BF03198064

Quintillan. (95AD). *The Institutio oratoria of Quintilian*. Cambridge, Mass. : Harvard University Press. Retrieved from http://www.archive.org/details/institutioorator00quin

Rasch, B., & Born, J. (2013). About Sleep's Role in Memory. *Physiological Reviews*, *93*(2), 681–766. doi:10.1152/physrev.00032.2012

Rasch, B., Büchel, C., Gais, S., & Born, J. (2007). Odor Cues During Slow-Wave Sleep Prompt Declarative Memory Consolidation. *Science*, *315*(5817), 1426 –1429. doi:10.1126/science.1138581

Rechtschaffen, A., & Kales, A. (1968). *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*. US Government Printing Office, US Public Health Service.

Rescorla, R. A. (2004). Spontaneous Recovery. *Learning & Memory*, *11*(5), 501–509. doi:10.1101/lm.77504

Reyna, V. F., & Brainerd, C. J. (1990). Fuzzy processing in transitivity development. *Annals of Operations Research*, *23*(1), 37–63. doi:10.1007/BF02204838

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*(1), 1–75. doi:10.1016/1041-6080(95)90031-4

Reyna, V. F., & Kiernan, B. (1994). Development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology*, *30*(2), 178–191. doi:10.1037/0012-1649.30.2.178

Ribeiro, S., Goyal, V., Mello, C. V., & Pavlides, C. (1999). Brain gene expression during REM sleep depends on prior waking experience. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *6*(5), 500–508.

Ribot, T. (1882). *Diseases of memory: an essay in the positive psychology*. D. Appleton and company.

Robertson, E. M. (2009). From Creation to Consolidation: A Novel Framework for Memory Processing. *PLoS Biol*, *7*(1), e1000019. doi:10.1371/journal.pbio.1000019

Robertson, E. M., Pascual-Leone, A., & Press, D. Z. (2004). Awareness Modifies the Skill-Learning Benefits of Sleep. *Current Biology*, *14*(3), 208–212. doi:10.1016/j.cub.2004.01.027

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. doi:10.1037/0278-7393.21.4.803

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: a parallel distributed processing approach*. MIT Press.

Rosanova, M., & Ulrich, D. (2005). Pattern-Specific Associative Long-Term Potentiation Induced by a Sleep Spindle-Related Spike Train. *The Journal of Neuroscience*, *25*(41), 9398–9405. doi:10.1523/JNEUROSCI.2149-05.2005

Rosenbaum, R. S., Winocur, G., & Moscovitch, M. (2001). New views on old memories: re-evaluating the role of the hippocampal complex. *Behavioural Brain Research*, *127*(1–2), 183–197. doi:10.1016/S0166-4328(01)00363-1

Roth, D. A.-E., Kishon-Rabin, L., Hildesheimer, M., & Karni, A. (2005). A latent consolidation phase in auditory identification learning: Time in the awake state is sufficient. *Learning & Memory*, *12*(2), 159–164. doi:10.1101/87505

Rothkopf, E. Z. (1957). A deduction from an excitation-inhibition account of retroactive inhibition. *Journal of Experimental Psychology*, *53*(3), 207–213. doi:10.1037/h0044851

Rubinsten, O., Henik, A., Berger, A., & Shahar-Shalev, S. (2002). The Development of Internal Representations of Magnitude and Their Association with Arabic Numerals. *Journal of Experimental Child Psychology*, *81*(1), 74–92. doi:10.1006/jecp.2001.2645

Rudoy, J. D., Voss, J. L., Westerberg, C. E., & Paller, K. A. (2009). Strengthening Individual Memories by Reactivating Them During Sleep. *Science*, *326*(5956), 1079–1079. doi:10.1126/science.1179013

Ryan, J. D., Moses, S. N., & Villate, C. (2009). Impaired relational organization of propositions, but intact transitive inference, in aging: Implications for understanding underlying neural integrity. *Neuropsychologia*, *47*(2), 338–353. doi:10.1016/j.neuropsychologia.2008.09.006

Ryan, L., Nadel, L., Keil, K., Putnam, K., Schnyer, D., Trouard, T., & Moscovitch, M. (2001). Hippocampal complex and retrieval of recent and very remote autobiographical memories: Evidence from functional magnetic resonance imaging in neurologically intact people. *Hippocampus*, *11*(6), 707–714. doi:10.1002/hipo.1086

Salasoo, A., Shiffrin, R. M., & Feustel, T. C. (1985). Building permanent memory codes: Codification and repetition effects in word identification. *Journal of Experimental Psychology: General*, *114*(1), 50–77. doi:10.1037/0096-3445.114.1.50

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. doi:10.1093/cercor/bhr034

Saletin, J. M., & Walker, M. P. (2012). Nocturnal mnemonics: sleep and hippocampal memory processing. *Frontiers in Neurology*, *3*, 59. doi:10.3389/fneur.2012.00059

Sara, S. J., & Hars, B. (2006). In memory of consolidation. *Learning & Memory*, *13*(5), 515–521. doi:10.1101/lm.338406

Schabus, M. (2009). Still Missing Some Significant Ingredients. *Sleep*, *32*(3), 291–293.

Schabus, M., Dang-Vu, T. T., Albouy, G., Balteau, E., Boly, M., Carrier, J., … Maquet, P. (2007). Hemodynamic cerebral correlates of sleep spindles during human non-rapid eye movement sleep. *Proceedings of the National Academy of Sciences*, *104*(32), 13164–13169. doi:10.1073/pnas.0703084104

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.

Schabus, M., Gruber, G., Parapatics, S., Sauter, C., Klösch, G., Anderer, P., … Zeitlhofer, J. (2008). Sleep spindles and their significance for declarative memory consolidation. *Sleep*, *27*(8), 1479–1485.

Schabus, M., Hödlmoser, K., Gruber, G., Sauter, C., Anderer, P., Klösch, G., … Zeitlhofer, J. (2006). Sleep spindle-related activity in the human EEG and its relation to general cognitive and learning abilities. *European Journal of Neuroscience*, *23*(7), 1738–1746. doi:10.1111/j.1460-9568.2006.04694.x

Schabus, M., Hödlmoser, K., Pecherstorfer, T., & Klösch, G. (2005). Influence of Midday Naps on Declarative Memory Performance and Motivation. *Somnologie*, *9*(3), 148–153. doi:10.1111/j.1439-054X.2005.00054.x

Schmidt, C., Peigneux, P., Muto, V., Schenkel, M., Knoblauch, V., Münch, M., … Cajochen, C. (2006). Encoding Difficulty Promotes Postlearning Changes in Sleep Spindle Activity during Napping. *The Journal of Neuroscience*, *26*(35), 8976 –8982. doi:10.1523/JNEUROSCI.2464-06.2006

Scoville, W. B., & Milner, B. (1957). Loss of Recent Memory After Bilateral Hippocampal Lesions. *J Neuropsychiatry Clin Neurosci*, *12*(1), 103–a–113. doi:10.1176/appi.neuropsych.12.1.103-a

Scullin, M. K. (2012). Sleep, Memory, and Aging: The Link Between Slow-Wave Sleep and Episodic Memory Changes From Younger to Older Adults. *Psychology and Aging*. doi:10.1037/a0028830

Shambroom, J. R., Fabgregas, S. E., & Johnstone, J. (2012). Validation of an automated wireless system to monitor sleep in healthy adults. *Journal of Sleep Research*. doi:10.1111/j.1365-2869.2011.00944.x

Shen, B., & McNaughton, B. L. (1996). Modeling the spontaneous reactivation of experience-specific hippocampal cell assembles during sleep. *Hippocampus*, *6*(6), 685–692. doi:10.1002/(SICI)1098-1063(1996)6:6<685::AID-HIPO11>3.0.CO;2-X

Siapas, A. G., & Wilson, M. A. (1998). Coordinated Interactions between Hippocampal Ripples and Cortical Spindles during Slow-Wave Sleep. *Neuron*, *21*(5), 1123–1128. doi:16/S0896-6273(00)80629-7

Siegel, J. M. (2001). The REM Sleep-Memory Consolidation Hypothesis. *Science*, *294*(5544), 1058–1063. doi:10.1126/science.1063049

Siemann, M., & Delius, J. D. (1996). Influences of task concreteness upon transitive responding in humans. *Psychological Research*, *59*(2), 81–93. doi:10.1007/BF01792429

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: JCSM: Official Publication of the American Academy of Sleep Medicine*, *3*(2), 121–131.

Sirota, A., & Buzsáki, G. (2005). Interaction between neocortical and hippocampal networks via slow oscillations. *Thalamus & Related Systems*, *3*(04), 245–259. doi:10.1017/S1472928807000258

Sirota, A., Csicsvari, J., Buhl, D., & Buzsáki, G. (2003). Communication between neocortex and hippocampus during sleep in rodents. *Proceedings of the National Academy of Sciences*, *100*(4), 2065 –2069. doi:10.1073/pnas.0437938100

Slamecka, N. J. (1966). Supplementary report: A search for spontaneous recovery of verbal associations. *Journal of Verbal Learning and Verbal Behavior*, *5*(2), 205–207. doi:16/S0022-5371(66)80019-1

Slamecka, N. J., & Ceraso, J. (1960). Retroactive and proactive inhibition of verbal learning. *Psychological Bulletin*, *57*(6), 449–475. doi:10.1037/h0048601

Smith, C., & Squire, L. R. (2005). Declarative Memory, Awareness, and Transitive Inference. *The Journal of Neuroscience*, *25*(44), 10138 –10146. doi:10.1523/JNEUROSCI.2731-05.2005

Smith, T. (2001). Sleep states and memory processes in humans: procedural versus declarative memory systems. *Sleep Medicine Reviews*, *5*(6), 491–506. doi:10.1053/smrv.2001.0164

Snowden, J. S., Griffiths, H. L., & Neary, D. (1996). Semantic-Episodic Memory Interactions in Semantic Dementia: Implications for Retrograde Memory Function. *Cognitive Neuropsychology*, *13*(8), 1101–1139. doi:10.1080/026432996381674

Song, S. (2009). Consciousness and the consolidation of motor learning. *Behavioural Brain Research*, *196*(2), 180–186. doi:10.1016/j.bbr.2008.09.034

Spencer, R. M. C., Sunm, M., & Ivry, R. B. (2006). Sleep-Dependent Consolidation of Contextual Learning. *Current Biology*, *16*(10), 1001–1005. doi:10.1016/j.cub.2006.03.094

Spriggs, W. H. (2008). *Essentials of Polysomnography* (1st ed.). Jones and Bartlett Publishers.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*(2), 195–231. doi:10.1037/0033-295X.99.2.195

Squire, L. R., Haist, F., & Shimamura, A. P. (1989). The neurology of memory: quantitative assessment of retrograde amnesia in two groups of amnesic patients. *The Journal of Neuroscience*, *9*(3), 828–839.

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.

Steriade, M. (2001). Active neocortical processes during quiescent sleep. *Archives Italiennes de Biologie*, *139*(1), 37–51. doi:10.4449/aib.v139i1.203

Steriade, M. (2006). Grouping of brain rhythms in corticothalamic systems. *Neuroscience*, *137*(4), 1087–1106. doi:10.1016/j.neuroscience.2005.10.029

Sterpenich, V., Albouy, G., Boly, M., Vandewalle, G., Darsaud, A., Balteau, E., … Maquet, P. (2007). Sleep-Related Hippocampo-Cortical Interplay during Emotional Memory Recollection. *PLoS Biol*, *5*(11), e282. doi:10.1371/journal.pbio.0050282

Sterpenich, V., Albouy, G., Darsaud, A., Schmidt, C., Vandewalle, G., Dang Vu, T. T., … Maquet, P. (2009). Sleep Promotes the Neural Reorganization of Remote Emotional Memory. *Journal of Neuroscience*, *29*(16), 5143–5152. doi:10.1523/JNEUROSCI.0561-09.2009

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

Stickgold, R. (2009). How Do I Remember? Let Me Count the Ways. *Sleep Medicine Reviews*, *13*(5), 305–308. doi:10.1016/j.smrv.2009.05.004

Stickgold, R. (2013). Parsing the role of sleep in memory processing. *Current Opinion in Neurobiology*. doi:10.1016/j.conb.2013.04.002

Stickgold, R., James, L., & Hobson, J. A. (2000). Visual discrimination learning requires sleep after training. *Nat Neurosci*, *3*(12), 1237–1238. doi:10.1038/81756

Stickgold, R., Scott, L., Rittenhouse, C., & Hobson, J. A. (1999). Sleep-Induced Changes in Associative Memory. *Journal of Cognitive Neuroscience*, *11*(2), 182–193. doi:10.1162/089892999563319

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

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

Straube, B. (2012). An overview of the neuro-cognitive processes involved in the encoding, consolidation, and retrieval of true and false memories. *Behavioral and Brain Functions*, *8*(1), 35. doi:10.1186/1744-9081-8-35

Takashima, A., Nieuwenhuis, I. L. C., Jensen, O., Talamini, L. M., Rijpkema, M., & Fernández, G. (2009). Shift from hippocampal to neocortical centered retrieval network with consolidation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(32), 10087–10093. doi:10.1523/JNEUROSCI.0799-09.2009

Takashima, A., Petersson, K. M., Rutters, F., Tendolkar, I., Jensen, O., Zwarts, M. J., … Fernández, G. (2006). Declarative memory consolidation in humans: A prospective functional magnetic resonance imaging study. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(3), 756 –761. doi:10.1073/pnas.0507774103

Tamminen, J. (2010). *Learning new words: Effects of meaning, memory consolidation, and sleep*. University of York., York.

Tamminen, J., Davis, M. H., Merkx, M., & Rastle, K. (2012). The role of memory consolidation in generalisation of new linguistic information. *Cognition*, *125*(1), 107–112. doi:10.1016/j.cognition.2012.06.014

Tamminen, J., & Gaskell, M. G. (2008). Newly learned spoken words show long-term lexical competition effects. *Quarterly Journal of Experimental Psychology*, *61*(3), 361–371. doi:10.1080/17470210701634545

Tamminen, J., Lambon Ralph, M. A., & Lewis, P. A. (2012). The role of sleep spindles and slow oscillations in integrating new information in semantic memory. *Submitted*.

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep Spindle Activity is Associated with the Integration of New Memories and Existing Knowledge. *J. Neurosci.*, *30*(43), 14356–14360. doi:10.1523/JNEUROSCI.3028-10.2010

Teng, E., & Squire, L. R. (1999). Memory for places learned long ago is intact after hippocampal damage. *Nature*, *400*(6745), 675–677. doi:10.1038/23276

Thase, M. E. (1998). Depression, sleep, and antidepressants. *The Journal of Clinical Psychiatry*, *59 Suppl 4*, 55–65.

Tilley, A. J., & Empson, J. A. C. (1978). REM sleep and memory consolidation. *Biological Psychology*, *6*(4), 293–300. doi:10.1016/0301-0511(78)90031-5

Tilley, A. J., & Warren, P. (1983). Retrieval from semantic memory at different times of day. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *9*(4), 718–724.

Tononi, G., & Cirelli, C. (2003). Sleep and synaptic homeostasis: a hypothesis. *Brain Research Bulletin*, *62*(2), 143–150.

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

Treves, A., & Rolls, E. T. (1992). Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus*, *2*(2), 189–199. doi:10.1002/hipo.450020209

Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, *4*(3), 374–391. doi:10.1002/hipo.450040319

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. doi:10.1126/science.1135935

Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., … Morris, R. G. M. (2011). Schema-Dependent Gene Activation and Memory Encoding in Neocortex. *Science*, *333*(6044), 891–895. doi:10.1126/science.1205274

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

Tucker, M. A., Hirota, Y., Wamsley, E. J., Lau, H., Chaklader, A., & Fishbein, W. (2006). A daytime nap containing solely non-REM sleep enhances declarative but not procedural memory. *Neurobiology of Learning and Memory*, *86*(2), 241–247. doi:10.1016/j.nlm.2006.03.005

Tulving, E. (1972). Episodic and Semantic Memory. In *Organization of Memory*. New York: Academic Press.

Tulving, E. (1985). *Elements of Episodic Memory*. Oxford University Press, USA.

Tulving, E. (2002). Episodic Memory: From Mind to Brain. *Annual Review of Psychology*, *53*(1), 1–25. doi:10.1146/annurev.psych.53.100901.135114

Underwood, B. J. (1948a). Retroactive and proactive inhibition after five and forty-eight hours. *Journal of Experimental Psychology*, *38*, 29–38.

Underwood, B. J. (1948b). ‘Spontaneous recovery’ of verbal associations. *Journal of Experimental Psychology*, *38*(4), 429–439. doi:10.1037/h0059565

Underwood, B. J., & Postman, L. (1960). Extraexperimental sources of interference in forgetting. *Psychological Review*, *67*(2), 73–95. doi:10.1037/h0041865

Valldeoriola, F., Santamaria, J., Graus, F., & Tolosa, E. (1993). Absence of REM sleep, altered NREM sleep and supranuclear horizontal gaze palsy caused by a lesion of the pontine tegmentum. *Sleep*, *16*(2), 184–188.

Van der Helm, E., Gujar, N., Nishida, M., & Walker, M. P. (2011). Sleep-Dependent Facilitation of Episodic Memory Details. *PLoS ONE*, *6*(11), e27421. doi:10.1371/journal.pone.0027421

Van Elzakker, M., O'Reilly, R. C., & Rudy, J. W. (2003). Transitivity, flexibility, conjunctive representations, and the hippocampus. I. An empirical analysis. *Hippocampus*, *13*(3), 334–340. doi:10.1002/hipo.10083

Van Kesteren, M. T. R., Fernández, G., Norris, D. G., & Hermans, E. J. (2010). Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. *Proceedings of the National Academy of Sciences*, *107*(16), 7550 –7555. doi:10.1073/pnas.0914892107

Van Kesteren, M. T. R., Rijpkema, M., Ruiter, D. J., & Fernández, G. (2010). Retrieval of Associative Information Congruent with Prior Knowledge Is Related to Increased Medial Prefrontal Activity and Connectivity. *The Journal of Neuroscience*, *30*(47), 15888–15894. doi:10.1523/JNEUROSCI.2674-10.2010

Van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*(4), 211–219. doi:10.1016/j.tins.2012.02.001

Van Ormer, E. B. (1932). Retention after intervals of sleep and waking. *Archives of Psychology*, *137*, 49.

Vertes, R. P. (2004). Memory consolidation in sleep; dream or reality. *Neuron*, *44*(1), 135–148. doi:10.1016/j.neuron.2004.08.034

Vertes, R. P., & Eastman, K. E. (2000a). REM sleep is not committed to memory. *Behavioral and Brain Sciences*, *23*(06), 1057–1063. doi:10.1017/S0140525X00984024

Vertes, R. P., & Eastman, K. E. (2000b). The Case Against Memory Consolidation in Rem Sleep. *Behavioral and Brain Sciences*, *23*(6), 867–876.

Vertes, R. P., & Siegel, J. M. (2005). Time for the sleep community to take a critical look at the purported role of sleep in memory processing. *Sleep*, *28*(10), 1228–1229; discussion 1230–1233.

Von Fersen, L., Wynne, C. D. L., Delius, J. D., & Staddon, J. E. R. (1991). Transitive Inference Formation in Pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *17*(3), 334–341.

Wagner, U., Gais, S., & Born, J. (2001). Emotional Memory Formation Is Enhanced across Sleep Intervals with High Amounts of Rapid Eye Movement Sleep. *Learning & Memory*, *8*(2), 112–119. doi:10.1101/lm.36801

Wagner, U., Gais, S., Haider, H., Verleger, R., & Born, J. (2004). Sleep inspires insight. *Nature*, *427*(6972), 352–355. doi:10.1038/nature02223

Wagner, U., Hallschmid, M., Rasch, B., & Born, J. (2006). Brief Sleep After Learning Keeps Emotional Memories Alive for Years. *Biological Psychiatry*, *60*(7), 788–790. doi:10.1016/j.biopsych.2006.03.061

Walker, M. P. (2005). A Refined Model of Sleep and the Time Course of Memory Formation. *Behavioral and Brain Sciences*, *28*(01), 51–64. doi:10.1017/S0140525X05000026

Walker, M. P., Brakefield, T., Allan Hobson, J., & Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature*, *425*(6958), 616–620. doi:10.1038/nature01930

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. doi:10.1016/S0896-6273(02)00746-8

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 & Memory*, *10*(4), 275–284. doi:10.1101/lm.58503

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. *Brain Research. Cognitive Brain Research*, *14*(3), 317–324.

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

Walker, M. P., & Stickgold, R. (2006). Sleep, Memory, and Plasticity. *Annual Review of Psychology*, *57*(1), 139–166. doi:10.1146/annurev.psych.56.091103.070307

Walker, M. P., & Stickgold, R. (2010). Overnight alchemy: sleep-dependent memory evolution. *Nature Reviews. Neuroscience*, *11*(3), 218; author reply 218. doi:10.1038/nrn2762-c1

Wamsley, E. J., Tucker, M. A., Payne, J. D., & Stickgold, R. (2010). A brief nap is beneficial for human route-learning: The role of navigation experience and EEG spectral power. *Learning & Memory*, *17*(7), 332–336. doi:10.1101/lm.1828310

Watts, A., Gritton, H. J., Sweigart, J., & Poe, G. R. (2012). Antidepressant Suppression of Non-REM Sleep Spindles and REM Sleep Impairs Hippocampus-Dependent Learning While Augmenting Striatum-Dependent Learning. *The Journal of Neuroscience*, *32*(39), 13411–13420. doi:10.1523/JNEUROSCI.0170-12.2012

Werchan, D. M., & Gómez, R. L. (2013). An interaction between reinforcement learning and sleep to facilitate transitive inference. *Neurobiology of Learning and Memory*, *100*, 70–76. doi:10.1016/j.nlm.2012.12.006

Werk, C. M., Harbour, V. L., & Chapman, C. A. (2005). Induction of long-term potentiation leads to increased reliability of evoked neocortical spindles in vivo. *Neuroscience*, *131*(4), 793–800. doi:10.1016/j.neuroscience.2004.12.020

Wilhelm, I., Diekelmann, S., Molzow, I., Ayoub, A., Mölle, M., & Born, J. (2011). Sleep Selectively Enhances Memory Expected to Be of Future Relevance. *The Journal of Neuroscience*, *31*(5), 1563 –1569. doi:10.1523/JNEUROSCI.3575-10.2011

Wilhelm, I., Rose, M., Imhof, K. I., Rasch, B., Büechel, C., & Born, J. (2013). The sleeping child outplays the adult's capacity to convert implicit into explicit knowledge. *Nature Neuroscience*, *advance online publication*. doi:10.1038/nn.3343

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

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

Wixted, J. T. (2004). The Psychology and Neuroscience of Forgetting. *Annual Review of Psychology*, *55*(1), 235–269. doi:10.1146/annurev.psych.55.090902.141555

Wright, B. [1], & Howells, D. [1]. (2008). Getting one step closer to deduction: Introducing an alternative paradigm for transitive inference. *Thinking & Reasoning*, *14*, 244–280. doi:10.1080/13546780802110131

Wright, P. K., Johnstone, J., Fabregas, S. E., & Shabroom, J. R. (2008). Evaluation of a portable dry sensor based automatic sleep monitoring system. *Sleep*, *31*, A337 – A337.

Wynne, C. (1995). Reinforcement accounts for transitive inference performance. *Learning & Behavior*, *23*(2), 207–217. doi:10.3758/BF03199936

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

Zeithamova, D., Schlichting, M. L., & Preston, A. R. (2012). The hippocampus and inferential reasoning: building memories to navigate future decisions. *Frontiers in Human Neuroscience*, *6*, 70. doi:10.3389/fnhum.2012.00070

Zeitlhofer, J., Gruber, G., Anderer, P., Asenbaum, S., Schimicek, P., & Saletu, B. (1997). Topographic distribution of sleep spindles in young healthy subjects. *Journal of Sleep Research*, *6*(3), 149–155. doi:10.1046/j.1365-2869.1997.00046.x

Zola-Morgan, S. M., & Squire, L. R. (1990). The primate hippocampal formation: evidence for a time-limited role in memory storage. *Science*, *250*(4978), 288–290. doi:10.1126/science.2218534

Zola-Morgan, S. M., & Squire, L. R. (1993). Neuroanatomy of Memory. *Annual Review of Neuroscience*, *16*(1), 547–563. doi:10.1146/annurev.ne.16.030193.002555