Effects of Sleep on Memory Consolidation and Automaticity in Word Learning

Elaine Kwang Hsia Tham

Doctor of Philosophy

University of York

Psychology

February 2014

Abstract

Compared to sleep studies on memory consolidation, there have been limited studies investigating effects of sleep on memory integration. Memory integration occurs when long-term associations between new and existing items are formed, leading to generalization of item knowledge across new contexts. Using semantic distance and size congruity effects as measures of automatic semantic access in integrated word representations, this thesis aimed to investigate the effects of sleep on declarative knowledge integration. By examining relationships between behavioural data and sleep architecture, this thesis also aimed to investigate if sleep plays an active role in knowledge integration. Experiment 1 revealed that participants who slept after learning exhibited significantly stronger semantic distance and size congruity effects for new second language words compared to participants who remained awake. Analysis of sleep polysomnography data also revealed that the strength of semantic distance and congruity effects were positively correlated with slow-wave sleep and sleep spindles respectively. Experiments 2 and 3 examined if selective factors such as training performance and encoding strength would modulate effects of sleep-dependent memory integration. Experiment 2 found that nap-associated gains in automaticity were stronger in participants who performed poorly during training. Experiment 3 also revealed that daytime-napping selectively benefitted integration of items that were more difficult to process. Experiment 4 explored the time-course required for automatic semantic access to emerge. Results highlighted that semantic distance effects emerge rapidly, whereas, size congruity effects that were considered a purer measure of automaticity only emerged after overnight consolidation. Finally, Experiment 5 studied if the sleep-associated findings from Experiments 1-4 would be similar for numeral learning. Compared to word learning, integration of new numerals required a shorter time-course and lower levels of sleep dependence. The thesis findings were evaluated against models of memory consolidation to gain theoretical understanding on the neural changes relating to knowledge integration.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my supervisor Gareth Gaskell for his support and mentorship throughout these years, and inspiring me to find joy in the research I carry out. I would like to express my gratitude to the members of my thesis advisory panel, Andy Ellis and Silke Göbel, for providing valuable feedback and useful insights to my research. I am also very grateful to the members of the sleep, language and memory (SLAM) laboratory and members of the psycholinguistics research group, who have been wonderful colleagues and treasured friends.

There are many friends who have supported me throughout my PhD journey. Although it is not possible to list everyone in this acknowledgement, I really want to say that I am truly thankful and feel richly blessed to have their friendship.

The greatest thanks must be extended to my fiancé Rayner and my family, who have accommodated my frequent discussions about sleep research, my irregular work hours and have often stayed up late to skype with me. I would not have been able to complete my PhD journey without their unconditional love and encouragement. I would like to say a special thanks to my mother, who has shared my joy in the good times and reassured me when I struggled – thank you for believing in me.

Declaration

This thesis contains original work completed solely by the author under the supervision of Professor Gareth Gaskell, with the exception of Experiment 3. Experiment 3 was conducted in conjunction with a final year undergraduate project that was supervised by Professor Gareth Gaskell and the author. Half of the experimental stimuli were generated by the undergraduate students and the students also ran approximately half of the participants.

The research was supported by funding from the Department of Psychology (University of York, UK) and the International Office (University of York, UK).

Data from this thesis were presented at the following conferences:

Experiment 1:
Tham, E.K.H., Lindsay, S. & Gaskell, M.G. *Effects of Sleep on Knowledge Integration and Automaticity of Processing*. Talk presented at the Experimental Psychology Society January meeting, London, UK, 2012

Experiment 1 and 2:
Tham, E.K.H., & Gaskell, M.G. *Sleep and Automaticity in Second Language Learning*. Talk presented at British Psychological Society Cognitive Section Meeting, Glasgow, UK, 2012

Experiments 1 and 2:
Tham, E.K.H., & Gaskell, M.G. *Automaticity in New Word Learning: Effects of Salience, Slow-wave Sleep, Spindles and Training Levels*. Talk presented at The 7th Asian Sleep Research Society Congress, Taipei, Taiwan, 2012

Experiments 4:
Tham, E.K.H., & Gaskell, M.G. *Direct and Active effects of Overnight Sleep on Knowledge Integration in New Word Learning*. Talk presented at The Donders Discussions, Nijmegen, Netherlands, 2013

Experiment 5:

Tham, E.K.H., & Gaskell, M.G. Effects of Sleep on Knowledge Integration and Automaticity of Processing: word learning and numerical cognition. Poster presented at the European Sleep Research Society Congress, Paris, France, 2012.

Chapter 1: The Role of Sleep in Knowledge Acquisition and Integration

1.1 Introduction:

Developments in psychological research have led to advances in the study of human memory by supplementing adult behavioural studies with developmental and neuroimaging data. The topic of memory has also been studied in relation to animal models, emotion and aging. An area that has also generated interest in researchers is the association between memory and sleep. Along with its role in memory consolidation, it has also been suggested that sleep is crucial for integrating new memories with existing knowledge.

This review aims to briefly introduce a biological definition of sleep including information of basic sleep architecture (section 1.2), followed by an overview of literature on the role of sleep on memory consolidation (section 1.3) and knowledge integration (section 1.4). It will then seek to consider the selective factors that may modulate sleep-dependent memory consolidation (section 1.5) and put forward an evaluation of theoretical models on sleep-dependent memory consolidation (section 1.6). The review will conclude with a proposed outline and aims for this thesis (section 1.7).

1.2 What is sleep?

1.2.1 Biological definition of Sleep

Sleep is a natural behaviour that is characterized by a reduction in both sensory activity and receptiveness to stimuli. It is easily reversible, unlike a period of hibernation or coma (Cirelli & Tononi, 2008). In humans, sleep is generally divided into non-rapid eye movement (NREM) and rapid eye movement (REM) sleep, which occur as an ultradian cycle of approximately 90 minutes across the night (Peigneux, Laureys, Delbeuck, & Maquet, 2001), and is characterized by two fundamental processes. First, a homeostasis process regulates duration of sleep whereby surplus sleep leads to an ensuing decreased drive for sleep, whereas sleep deprivation results in an increased drive (Borbély & Achermann, 1999). Second, a circadian process regulates the timing of sleep (Borbély, Achermann, Trachsel, & Tobler, 1989).

1.2.2 Basic sleep architecture: Sleep Cycle, Sleep stages

Brain activity during sleep can be measured using electroencephalography (EEG) and reflected in terms of frequency bands. NREM sleep is split into four stages, ordered according to decreasing frequency in EEG wave activity and increasing depth of sleep, whereby propensity to arousal is greatest at stage 1 and lowest at stage 4 (Rechtschaffen & Kales, 1968).

Stage 1 NREM sleep is characterized by mixed EEG of 2-7 Hz, whereas Stage 2 sleep is classified by the occurrence of sleep spindles (11-15 Hz waves that last for at least 0.5 seconds) and less than 20% of low frequency waves (Rechtschaffen & Kales, 1968). It is important to note that even though sleep spindles are also present in the later stages of NREM, they occur predominantly in Stage 2 sleep. Stages 3 and 4 (deepest stages) of NREM sleep are commonly referred to jointly as 'slow-wave sleep' (SWS) due to the large amounts of low frequency Delta waves (1-4 Hz and <1 Hz respectively) predominating the EEG recordings (Iber, Ancoli-Israel, Quan, 2007).

In contrast to NREM sleep, REM sleep is characterized as mixed frequency waves with low voltages accompanied by the incidence of rapid eye movements (Silber et al., 2007). Even though the NREM-REM cycle occurs at fairly constant intervals, the distribution of sleep stages varies. For instance, across a typical night of sleep, SWS would dominate the first half of the night; and stage 2 NREM and REM sleep would dominate the second half (Carskadon & Dement, 2000).

1.3 The Role of Sleep in Declarative Memory Consolidation

The main proposition that sleep is beneficial for learning stems from studies on sleep and memory consolidation (Walker & Stickgold, 2004). Memory consolidation is the process where a memory representation that is unstable, becomes stable and qualitatively more resistant to forgetting over time. Memory consolidation is thus important for learning as it promotes long-term maintenance and enhancement of information one acquires (Stickgold & Walker, 2005). Several behavioural, neuroimaging, molecular and animal studies have indicated that sleep benefits memory consolidation (Diekelmann & Born, 2010a; Stickgold, 2005; Walker & Stickgold, 2004).

Memory is often broadly categorised into non-declarative and declarative memory (Squire & Zola, 1996). Non-declarative memory comprises memories for procedural skills or habits, priming, conditioned responses and non-associative learning (Squire, 2011). Declarative memory comprises memory for specific events (episodic memory) or explicit factual knowledge (semantic memory) (Squire, Knowlton, & Musen, 1993; Squire & Zola, 1996). Even though sleep has been argued to affect both non-declarative and declarative memory, the scope of this review will focus on the effects of sleep on memory consolidation of declarative memory. In the recent decade, there has been an increase in the number of studies investigating the role of sleep on declarative memory consolidation. Still, the precise function of sleep on declarative memory consolidation remains widely debated. Some researchers have questioned whether sleep is even essential for memory consolidation. Amongst research on sleep dependent memory consolidation, some studies predict a passive role of sleep whereas others have proposed that sleep plays an active role for memory consolidation (for a more extensive review see Ellenbogen, Payne, & Stickgold, 2006; Mednick & Alaynick, 2010).

1.3.1 Sleep has no benefits towards declarative memory consolidation

Although not widely subscribed to amongst sleep researchers, some experimenters have suggested that sleep plays no beneficial role towards memory consolidation. In a review by Vertes & Eastman (2000) on antidepressant drugs that lead to REM sleep deprivation, the researchers concluded that the above drugs did not negatively affect declarative memory, hence sleep does not play a role in declarative memory consolidation. However, it is useful to note that most evidence against the role of sleep on declarative memory consolidation has focused on REM sleep (see Vertes, 2004), whereas sleep consists of both REM and NREM sleep. It is plausible that NREM sleep may be more involved in declarative memory consolidation than REM sleep or that REM sleep alone cannot account for declarative memory consolidation (this will be covered in section 1.5 of this review). Therefore, researchers may need to refer to the full sleep cycle consisting of both NREM and REM sleep when investigating the role of sleep on memory consolidation.

1.3.2 Sleep plays a passive role in declarative memory consolidation

Researchers who attribute a passive role of sleep in declarative memory consolidation have also focused on behavioural studies depicting that recall performance after sleep is significantly better than after an equivalent period of wakefulness. One potential interpretation is that such findings do not directly provide evidence that sleep benefits memory but rather implies that sleep shields new memories from retroactive interference (Ellenbogen, Payne, et al., 2006). One of the first studies that implied that sleep protects declarative memories from interference was conducted by Ellenbogen, Hulbert, Stickgold, Dinges, and Thompson-Schill, (2006) who looked at memory for trained word-pairs. Participants were train on a set of 'A-B' word-pairs in the evening (sleep group) or morning (wake group) and either underwent sleep or remained awake for a comparable duration after learning. Prior to testing, participants also experienced interference of a new 'A-C' pair list. Results indicated that participants who underwent sleep after learning 'A-B' word-pair recalled significantly more 'B'-words when tested after 12-hours, compared to participants who remained awake. It is therefore possible that sleep passively promotes the resistance of declarative memories to retroactive interference.

Brown and Lewandowsky (2010) argued that evidence for the benefit of sleep in preventing forgetting of learnt information due to interference, can be explained alternatively by temporal distinctiveness models. Temporal distinctiveness models predict that the longer the temporal space between learning of new information and subsequent interference, the weaker the effect of the interference on subsequent memory (Glenberg & Swanson, 1986). In relation to studies which found that sleeping soon after learning leads to better memory consolidation of learnt information (Lau, Alger, & Fishbein, 2011; Payne, Chambers, & Kensinger, 2012), Brown and Lewandowsky (2010) proposed that enhanced memory in the condition when participants slept soon after learning could be because of the longer the temporal space between learning of new information and subsequent interference due to the occurrence of sleep. In addition, Brown, Neath and Chater (2007) proposed a temporal distinctiveness model termed the scale-independent memory, perception and learning (SIMPLE), which predicted that greater the ratio of time between learning and retrieval the more fragile a memory was to interference. The SIMPLE model assumes that memories are influenced by many mechanisms in addition to temporal information; hence the model also predicts that newly acquired memory would be more dependent on temporal mechanism during retrieval as compared to remote memories, which may be more influenced by other mechanisms such as serial order (see Lewandowsky, Ecker, Farrell, & Brown, 2012 for an indepth description of the SIMPLE model). Therefore, applying the SIMPLE model to a simulation of retrograde amnesia, Lewandowsky, Ecker, Farrell and Brown (2012) suggested that classic studies such as the impairment of recent memories in retrograde amnesia cited as support for sleep-dependent consolidation, could be accounted for instead by an impairment in temporal representation of information instead.

1.3.3 Sleep plays an active role in declarative memory consolidation

However, apart from passive protection, sleep has been hypothesized to play a direct and active role in memory consolidation. Although memory consolidation can occur during wakefulness, it is mainly dependent on sleep. In the word-pair association study (Ellenbogen, Hulbert, et al., 2006) mentioned in the previous section 1.3.2, apart from passively shielding the initial 'A-B' word-pair from retroactive ('A-C' word-pair) interference, it is also feasible that sleep may actively strengthen the association of the initial 'A-B' pairing. In addition to the 12-hour wake and 12-hour sleep group, Ellenbogen et al. (2006) also tested a 24-hour group where participants were trained in the morning and tested 24 hours later. Even though participants in the 24-hour group experienced a night of sleep after testing, they first experienced a similar amount of wakeful activity as the 12-hour wake group. Results indicated that participants in the 24-hour group recalled significantly more 'B'-words than those in the wake group. This suggests that not only does sleep protect words from retroactive interference; it may actively strengthen the memory trace of learnt items. Despite experiencing a longer duration between training and testing than the wake group, the 24-hour group still performed better than the wake group. In addition, there was no significant difference in performance between the 24-hour and sleep group; suggesting that sleep rather than a passage of time is essential for consolidation of learnt items.

Moreover, research relating individual components of sleep architecture such as slow-wave activity (including slow-wave sleep) and sleep spindles to sleep associated behavioural changes provide evidence for the active role of sleep in benefitting declarative memory consolidation. The unique role of slow-wave sleep (SWS) on declarative memory consolidation stems from studies highlighting that hippocampal (declarative) memories in rodents were 'reactivated' during sleep (Ji & Wilson, 2007; Pavlides & Winson, 1989; Wilson & McNaughton, 1994). Wilson and McNaughton (1994) found that hippocampal place cells that only fire when the rats were in certain places (during wakefulness) also displayed a similar order of firing activity when the rats entered SWS. Most importantly, this reactivation was not present in any other sleep stage except in SWS.

Following the earlier studies on reactivation in animals, there has been an increase in recent research on memory reactivation in humans. Rasch, Büchel, Gais and Born (2007) trained participants in a declarative task where they had to associate specific odours with individual locations. The experimenters then reactivated participants' memories by presenting the odours during SWS, REM sleep or wakefulness. Participants who were presented with the odours during SWS had significantly better memory in the location recall as compared to a control condition. In contrast, there was no significant benefit in location recall memory compared to the control condition when the odours were reactivated during REM sleep or wakefulness. Moreover, the presence of odours during also SWS lead to a significant increase in hippocampal activity. The above behavioural and neuropsychological findings suggest that reactivations during SWS play a direct and active role in declarative memory consolidation (Rasch & Born, 2007). Further support for the active role of SWS on declarative memory consolidation was presented in a subsequent nap study by Rudoy, Voss, Westerberg and Paller (2009), where participants first learnt object-location pairs that were presented with semantically related sounds. The experimenters then presented half of the semantically related sounds to the participants during SWS. When participants were tested on their memory of the object-locations pairs after sleep, results revealed that participants performed with less forgetting for pairs that had their corresponding sounds replayed during SWS.

In addition to reactivation studies, evidence for the active role of sleep in declarative memory consolidation has also been drawn from external manipulation and enhancement of slow wave activity during sleep. It has been observed that cholinergic tone (based on the amount of the neurotransmitter acetylcholine) during SWS is lower than that of wakefulness or any other sleep stage. Therefore, Gais & Born (2004) manipulated the level of cholinergic tone in human participants during sleep by giving participants a dose of a cholinesterase inhibitor (physostigmine) before sleep to investigate its effect on declarative memory consolidation. The dose of physostigmine led to an increase in cholinergic levels during early sleep which mainly consists of SWS. The experimenters found that increasing cholinergic tone during SWS dominated sleep led to impaired performance on a declarative memory task but did not affect performance on a procedural task. The experimenters interpreted above outcome as an indication that the naturally occurring low cholinergic tone in SWS may be crucial for declarative memory consolidation, implicating an active role for SWS in declarative memory consolidation.

However, an alternative explanation can also be that cholinergic tone, rather than SWS is important for active memory consolidation. In order to attribute a more direct and active role of SWS to memory consolidation, it is useful to consider other techniques where external manipulation and enhancement of slow wave activity is achieved. On two separate sessions, Marshall, Helgadóttir, Mölle and Born (2006) applied transcranial slow oscillation stimulation or sham stimulation in participants' NREM after they were trained on a declarative word-pair associate task and a non-declarative finger tapping task. The transcranial slow oscillation stimulation led to an increase in 0.5-1.0 Hz EEG activity that paralleled markers of SWS. Performance in a post-sleep test indicated that participants were significantly better in the declarative word-pair association task after the slow oscillation stimulation as compared to the sham stimulation. In contrast, there was no difference between the effects of slow oscillation and sham stimulation in participants' performance for the non-declarative finger tapping task. Therefore, rather than cholinergic tone, SWS seems to be the factor that plays a direct role in benefitting declarative memory consolidation.

The relationship between sleep spindles and declarative performance also lends support to an active role of sleep in declarative memory consolidation. Gais, Mölle, Helms, and Born (2002) found that increased spindle density in post-training sleep was positively correlated to recall performance on a word-pair association task. Moreover, post-training sleep spindle density was also significantly higher after participants were trained on the declarative word-pair association task as compared to a control task which did not involve declarative learning. Hence, research on sleep spindle activity has also provided insight on the direct role for sleep in memory consolidation. In a recent clinical study, Mednick et al. (2013) also demonstrated that increased sleep spindle activity from intake of pharmacological drugs lead to enhanced performance in a declarative word-pair association task but not for a non-declarative motor task. Furthermore, Marshall et al. (2006) found that apart from an increase in 0.5-1.0 Hz EEG activity, transcranial slow oscillation stimulation during NREM sleep also lead to an increase in participants' sleep spindle activity such as spindle frequency and spindle count. For that reason, experimenters also predicted that the combination of slow-wave and sleep spindle activity plays a causal role in declarative memory consolidation. In summary, the direct relationships between SWS and sleep spindles on declarative memory suggest that sleep plays a direct role in declarative memory consolidation. The strong possibility that an interaction between slow-wave activity and sleep spindle activity (Marshall et al., 2006) leads to enhanced declarative memory performance further predicts that the role which sleep plays is not a passive role but rather an active and potentially causal one (see also systems consolidation model in section 1.5.2).

1.4 Sleep and Knowledge Integration

When contemplating the effects on sleep on learning and memory, it is also useful and important to consider the integration of new memories with existing knowledge. Researchers have often associated declarative memory consolidation to that for individual item memories which are mainly episodic in nature. However, most declarative learning involves assessing semantic information or meanings from items that ranges beyond episodic memory. Therefore, knowledge integration, whereby one forms long term associations between new item memories and existing information, leading to the generalization of information between items across new contexts; is also essential for declarative learning (Walker, 2009). Also, unlike quick initial (episodic) acquisition of the form of novel information, knowledge integration requires a longer time course and hence may also be paramount in reflecting neurological restructuring in the adult brain (Karni & Giuseppe, 1997).

There have been limited studies investigating the effects of sleep on memory integration compared to studies on general memory consolidation. Key studies in this area include work on human relational memory by Ellenbogen, Hu, Payne, Titone, and Walker (2007) where experimenters taught participants 5 separate ‘premise pairs’ (A>B...E>F) with a hidden hierarchy. Although participants could remember the relationship between the individual pairs after a 12 hour delay, only participants who slept after learning were able to make novel inferences between pairs separated by 2 units, such as B>E. This suggests that sleep enabled integration of information about individual pairs, allowing participants to gain insight to the generalized hierarchy that A>B>C>D>E.

In another relational memory study, participants learnt pairs of face-object associations during training, where each object was paired with two different faces on separate trials. During testing, in addition to being examined on the above pairings, participants also made judgments on relational face-face associations (deciding which two faces where paired with the same object) that were not explicitly learnt during training (Lau, Tucker, & Fishbein, 2010). Results indicated that participants who napped between training and testing performed significantly better on both direct face-object pairs and relational face-face associations as compared to participants who remained awake. Moreover, performance on the relational associations was significantly correlated with amount of SWS, whereas performance in the direct pairings was not related to any sleep parameter. The above findings highlight that sleep, particularly SWS, enhances integration of relational information between separately learnt items.

Still research on the integration of new knowledge into existing schemas has suggested that consolidation can occur speedily offline (see van Kesteren, Ruiter, Fernández, & Henson, 2012 for a review). Tse et al. (2007) trained rodents on a flavour-location paired-associate task. The task was typical of a hippocampal-dependent declarative task as rodents with hippocampal lesions only performed at chance level despite extensive training. On another group of rodents which were well-trained on the task, Tse et al. (2007) demonstrated that the rodents were able to learn a new set of flavour-location paired-associates in one trial without need for any offline consolidation. In addition, rodents which had their hippocampus removed after initial pair-associate learning were still able to learn the new flavour-location paired-associates. The experimenters emphasized that findings provided evidence for the 'schema hypothesis', when a new schema is consistent with an existing schema, information can be integrated rapidly. Tse et al. (2011) also provided addition support for the schema hypothesis by highlighting associated between schema learning and activation of genes in the neocortex. However it is important to note that Tse et al.'s (2007, 2011) findings does not show that all knowledge integration does not reply on offline consolidation. Rather, Tse et al. (2007, 2011) only predict that integrated of new schemas that are highly related to existing knowledge may be independent of offline consolidation (for more information see section 1.5.4).

In a recent relational memory study by Lau, Alger and Fishbein (2011), participants learnt 7 sets of Mandarin characters and their relevant English meanings. Each set of characters contained a common property known as radicals, which are often used to inform Mandarin readers about the meaning of the characters. After either a 90 minute nap or an equivalent period of wakefulness, participants were given tasks containing untrained Mandarin characters with the same radicals as the sets of trained characters. For each untrained Mandarin character, there 4 possible options as to what its correct meaning was. In order to get the correct answer for this task, participants had to abstract commonalities between the untrained character and the corresponding set of trained characters with the same radical. For the next task, participants were also shown individual radicals and had to indicate what they thought the radicals meant. Participants who had taken a nap between training and test achieved significantly higher accuracy compared to those who remained awake for both tasks. Hence, findings indicate that sleep is central in integrating information across meanings of new information, enabling abstraction of previously untrained semantic rules between the Mandarin characters.

Further evidence for the effect of sleep on knowledge integration comes from a body of work on novel word learning. Dumay & Gaskell (2007) also provided evidence for the benefit of sleep in knowledge integration in a study where they explored the effects of sleep on learning novel (spoken) words via a lexical competition paradigm; this occurs when a novel word (e.g. “cathedruke”) increases reaction time for an existing word (e.g. “cathedral). The experimenters suggested that this only occurs when a novel word has integrated into existing neocortical knowledge; causing it is be more quickly assessed and hence interfering with the time for recognizing an existing word (Gaskell & Dumay, 2003). Two groups of participants were tested three times: immediately, 12-hours and 24-hours after learning the novel words. The PM group learnt the words at night and slept before the 12-hours test; while the AM group learnt the words in the morning and remained awake till the 12-hours test. The lexical competition effect was present in the PM group for both tests but only present in the test conducted after 24-hours for the AM group, after participants slept. These findings provided support for the prediction that the process of integration is dependent on sleep rather than a general passage of time. In a subsequent study, Davis, Di Betta, Macdonald, & Gaskell (2009) trained participants on two sets of novel words over consecutive days (day 1/day 2). Tests were conducted on day 2, including lexical decision and recognition tasks. Participants performed significantly better on words learnt on day-1 compared to day 2 for all tasks. As day 1 but not day 2 words experienced overnight consolidation; participants’ performance supported findings by Dumay and Gaskell (2007), indicating that sleep enhances both consolidation of the individual novel words and integration of the words with existing lexical knowledge.

Tamminen, Payne, Stickgold, Wamsley and Gaskell (2010) suggested that different components of sleep may play unique roles in memory consolidation for novel words. Participants were trained on novel spoken words that were derived from similar sounding English base words used in previous consolidation experiments. Participants who slept after training experienced a greater lexical decision effect compared to those who remained awake for an equivalent period after training. Moreover, SWS was related to the strengthening of individual word memories as participants who experienced more SWS also showed greater improvements in speed of categorisation as to whether a word was a previously trained word or a foil. The study also used a lexical competition effect (Gaskell & Dumay, 2003) as an index of how well a novel word had been integrated into the lexicon. The size of this integration measure was correlated with spindle activity, highlighting that sleep spindle activity plays a crucial role in integrating novel words with existing lexical knowledge.

In summary, current research on relational memory and also on novel word learning give credence to the notion that sleep benefits knowledge integration. It is important to note that there may be different types of integration that are not often dissociated from each other. The integration mentioned in the relational memory studies (Ellenbogen et al., 2007; Lau et al., 2011) refer to an integration of associative concepts. In contrast, the integration mentioned in the word learning studies expands beyond the associative concepts and refer more to the integration of new word forms with existing neocortical mappings. Even though both types of integration have been broadly referred to as a single term in the sleep literature, they are qualitatively very different. This thesis will focus on the integration of newly learnt information and existing knowledge that occurs in novel word learning. Findings demonstrating that novel words experienced lexical competition with existing words after participants underwent sleep provide evidence which suggest that sleep may enhance integration of novel words with existing word-meaning mappings. Further evidence using sleep polysomnography also predict that sleep architecture such as sleep spindles may play a unique role towards knowledge integration.

1.5 Selective factors that modulate sleep-dependent consolidation

There has been a fairly large selection of research indicating that sleep benefits consolidation of newly learnt items (Diekelmann & Born, 2010a; Rasch &

Born, 2013). However, there has been limited work investigating if sleep benefits all learning equally or if sleep selectively benefits learning of certain items more than others. This review will focus on the differences in encoding strength (performance during initial learning) and also look at the effects of item saliency when considering factors that may modulate sleep-dependent consolidation.

1.5.1 Effects of encoding strength

Current research regarding effects of strength of initial encoding and sleep-dependent consolidation has yielded conflicting results. An overnight study by Drosopoulos, Schulze, Fischer and Born (2007) found that following retroactive interference from a new set of word-pair associates, performance on initial weakly encoded word-pairs was better after sleep compared to an equivalent period of wakefulness. Moreover, there were no differences between wake and sleep groups for intensely encoded word-pairs, highlighting that the beneficiary effect of sleep against subsequent interference was unique to weakly encoded items. In contrast, from their study of face-location associative memory, Talamini, Nieuwenhuis, Takashima, and Jensen (2008) highlighted that memories that were strongly encoded at the time of overnight sleep onset were significantly more resistance to forgetting than weakly encoded memories. Tucker and Fishbein (2008) also found that nap participants who performed well during initial learning of unrelated word-pairs, recalled significantly more items than participants who remained awake. Additionally, when looking at participants who performed poorly during learning, there were no significant differences between wake and nap groups on post-learning recall. Still, in another study where encoding difficulty was manipulated in a within-subjects design, experimenters found no difference in improvements relating to sleep-dependent consolidation between 'easy-encoding' and 'difficult-encoding' word-pairs (Schmidt et al., 2006).

Furthermore, similar to the above behavioural findings, the relationship between individual sleep components, differences in learning and post-sleep performance is also not conclusive. Drosopoulos et al. (2007) did not report any relationships between sleep parameters and weak or intense encoding strength. In contrast, Tucker and Fishbein (2008) found that SWS was positively related to post-training improvements in levels of word-pair recall for participants who performed well during the learning phase, but there was no effect of SWS for participants who performed poorly during the learning phase. However, Schmidt et al. (2006) did not find any relationship between encoding difficulty during learning to EEG power and

SWS, but found that sleep-dependent improvements for 'difficult encoding' items were positively correlated with low frequency spindle activity in the left frontal EEG channel.

1.5.2 Effects of item saliency

Next, this review will also consider the effects of item saliency in modulating sleep-dependent memory consolidation. Saliency can be broadly defined as a factor that leads to increased cognitive awareness such that the item or thought stands out (Myers & Alpert, 1977). Hu, Stylos-Allan, and Walker (2006) found that at a test session after 12-hours of viewing a set of emotionally arousing pictures and neutral pictures, participants recalled significantly more emotionally arousing stimuli when the 12-hour interval included a night of sleep as compared to wakefulness. In contrast, there was no effect of sleep over wakefulness for the neutral pictures, suggesting that sleep may selectively benefit emotionally salient items. Payne, Chambers, and Kensinger (2012) provided additional support towards the selective benefit of sleep for emotionally salient items when they revealed that for participants who experienced post-learning overnight sleep on the same day of learning, the enhancement in recognition of emotionally salient aspects of scenes was related to worse recognition of emotionally neutral aspects. It is important to note that the investigation of the effects of item saliency on sleep-dependent memory consolidation has mainly been conducted on emotional memories.

However, based on Myers and Alperts' definition, saliency can also manifest in items that are not manipulated for emotionality. In addition to emotional salience, Saletin and Walker (2012) proposed that prefentical sleep-associated consolidation of salient items could also include saliency in terms of originality, future incentive and level of explicitness during training. Wilhelm et al. (2011) trained participants in a delcarative word-pair association task where half the participants were told to expect a future memory test and the other half were not. All participants were also test after a period of overnight sleep or wakefulness. The experimenters found that participants who expected the memory test and slept after learning recalled significantly more word-pairs than participants who did not expect the memory test and also participants who expected the test but remained awake after learning. In addition, amongst participants who slept, those who expected the post-sleep memory test had significantly more SWS than those who did not expect the test. Instead of manipulating future relevance in a between-subjects way, van Dongen, Thielen, Takashima, Barth and Fernández (2012) trained participants in a picture-location paired-associate task where participants were told that only half the paired-associates would be tested. Van Dongen et al.'s (2012) within-subject manipulation of relevance replicated Wilhelm et al.'s (2011) finding as participants who slept after learning only showed significant enhance memory compared to those who remained awake for the relevant picture-location pairs. There was no significant benefit of sleep for picture-location pairs that participants thought would not be tested. In summary, sleep does not seem to benefit all dclarative memories equally. Current findings suggest that sleep preferentially benefits memories that are salient, which includes emotional memories and memories that that relevant for the future.

1.6 Evaluation of sleep-dependent consolidation models

Across sleep literature, researchers have subscribed to different theoretical models in order to best explain findings on the effects of sleep on memory consolidation. Although the list is not exhaustive, this review will discuss some key models used by sleep researchers to gain insights on the effects of sleep in memory consolidation. As there has been a lack of theoretical models directly explaining the effects of sleep in knowledge integration, this review will also consider how the models can be potentially adapted to explain integration.

1.6.1 Dual process hypothesis model

Many early sleep researchers who looked at effects of individual sleep stages on memory consolidation have often attributed their findings to the dual process hypothesis model. The model hypothesizes a separate role for REM and NREM sleep on different categories of memories; whereby REM sleep and SWS uniquely benefit consolidation of procedural and declarative memories respectively (Peigneux et al., 2001). Evidence supporting the above model was first derived from dissociations in a seminal study by Plihal & Born (1997) where experimenters found that early sleep (comprising mainly of SWS) improved recall for a declarative word-pair association task whereas late sleep (comprising longer durations of REM than early sleep) improved recall on a procedural mirror tracing task. In addition, a subsequent study by Tucker et al., (2006), where participants underwent a daytime nap (approximately 48% SWS but no REM sleep) or an equivalent period of wakefulness revealed no group differences in performance on a similar procedural mirror tracing task. In contrast, participants in the nap group showed a significantly larger improvement for the word-pair association task compared to those who remained awake.

Still, when referring to NREM sleep, the above studies did not look at the effects of Stage 1 and Stage 2 NREM sleep separately from that of SWS and hence, were not sufficient to act as primary evidence for a causal role linking REM sleep and SWS to the benefit of procedural and declarative memory consolidation respectively. Additionally, results from other studies which contradict the claims made by the dual process hypothesis model, have also raised doubts as to whether the model is sufficient to explain all incidences of sleep-dependent memory consolidation. Gais, Plihal, Wagner and Born (2000) found that early sleep consisting largely on SWS facilitated memory for a procedural discrimination task. In a separate study, researchers found a significant increase in REM theta power after participants completed a declarative task (Fogel, Smith, & Cote, 2007). Therefore, these examples highlight that the role for REM and NREM sleep in memory consolidation may not be mutually exclusive. Hence, it is feasible that sleep stages have a complementary role in consolidation of all forms of memories, with certain sleep stages showing greater enhancement in particular categories of memory than others. For example, although SWS may be essential for the consolidation of declarative memory, this does not discount a potentially facilitative role REM sleep may also have in declarative memory consolidation.

1.6.2 General two-stage systems consolidation model

The two-stage systems consolidation model seeks to address the key question relating to the process of how newly learnt short-term declarative memories are consolidated into long-term memories. The model encompasses two neural systems: the hippocampal system and the neocortical system; where the hippocampal system acts as a short-term store for new and distinct declarative memories. In contrast, the neocortical system serves as a permanent store for long-term memories (Diekelmann & Born, 2010a; Frankland & Bontempi, 2005). Support for the model comes from the notion of brain plasticity coupled with the widely held idea that memory is not a singular concept (Squire, 2004; Squire & Zola, 1996). A key example arises from a seminal study in neuroscience of patient HM who developed amnesia after a bilateral hippocampal lesions (Scoville & Milner, 1957). HM’s memory loss was specific to recent memories (anterograde amnesia) as he was able to recall events prior to the lesions. Moreover, recent reviews on HM memory loss revealed that although he was impaired in lexical decision and novel word completion learning tasks, he was able to make improvements in other visuo-spatial learning tasks such as mirror tracing (see review by Corkin, 2002). Hence, the bilateral hippocampal lesions seemed to cause a specific impairment in HM's ability to acquire new declarative memory despite preserving his long-term memory and memory for non-declarative learning. This directly implies that more than one neural system is needed for consolidation of new memories.

Unlike the Dual Process Hypothesis, the general two-stage systems consolidation model puts forward a mutually inclusive role for both NREM and REM sleep where both complement each other in their role towards memory consolidation. Buzsáki (1989) who was one of the pioneers in developing the model, hypothesized that the sharp waves which occur mainly in SWS enabled the transfer to information from the hippocampus to neocortex through the firing of synchronous neocortical neurons. In contrast, increased theta waves during REM sleep leads to an opposite neocortex to hippocampal transfer route whereby the above on-going 'dialogue' between both hippocampal and neocortical systems contribute to the consolidation of new memories (Buzsáki, 1996). According to adaptations of the system consolidation model, novel declarative information is first encoded and represented in both the hippocampal and neocortical systems. During SWS, the on-going 'dialogue' between both systems, and the reactivation of hippocampal memories (see section 1.3.3. sleep plays an active role in declarative memory consolidation) leads to a transfer of representations in the hippocampal system to the neocortical system. The repetition of the above process across multiple nights of sleep enables the novel memories to be independent of the hippocampal systems and fully dependent on the neocortical system for retrieval (Diekelmann & Born, 2010a, 2010b).

1.6.3 Sequential hypothesis model

Drawing from the systems consolidation model, the Sequential hypothesis model proposes a sequential complementary role of SWS and REM in memory consolidation, where both sleep stages are not independent of each other and that the memory trace stored during wakefulness has to be firstly modulated during SWS and subsequently REM sleep to be efficiently consolidated in memory. The flow of events also mirrors the normal sleep cycle in adult mammals, where SWS precedes REM sleep (Giuditta et al., 1995).

Mednick, Nakayama and Stickgold (2003) investigated the effects of daytime napping on a procedural textile discrimination task where participants remained awake, took a nap with SWS (no REM sleep) or took a nap with both SWS and REM sleep after training. The experimenters found that naps containing SWS prevented deterioration of performance in the task compared to wakefulness. However, naps with both SWS and REM led to improvements in performance; supporting the Sequential hypothesis model that SWS and REM sleep are not independent in their role towards benefitting memory consolidation.

Next, studies examining disruptions in the natural NREM (SWS) - REM sleep cycle lends addition support to the sequential hypothesis model. As the average NREM-REM sleep cycle duration decreases with age, one study investigated the role of the human sleep cycle in memory consolidation using a word recall task in ageing participants (range: 63-73 years) (Mazzoni et al., 1999). Results denoted that performance in word recall was significantly positively correlated with average NREM-REM sleep cycle duration. Moreover, there was no significant relationship between recall and total time spent in NREM or REM sleep, highlighting the importance of the sleep cycle as a whole rather than individual sleep stages for memory consolidation.

In general, studies supporting the Sequential hypothesis model have been fairly successful in explaining general sleep-dependent gains in learning. However, there have been no known studies that clearly link the model with findings on the effects of sleep on knowledge integration.

1.6.4 Complementary learning systems model

Similar to general two-stage systems consolidation model, the complementary learning systems model (CLS) draws upon evidence supporting the presence of separate hippocampal and neocortical systems (McClelland, McNaughton, & O'Reilly, 1995) and incorporates the role of both REM and NREM sleep in memory consolidation. The model draws from connectionist theories and provides a potential solution to the issue of how the brain learns large amounts of novel declarative information and incorporates them into the neocortical system without causing interference to the system (McClelland & Goddard, 1996; McClelland et al., 1995).

By adapting connectionist modelling with systems consolidation, the CLS proposes that the hippocampal system acts as a fast learning store that encodes distinct declarative information rapidly in contrast to the neocortical systems that uses a slower learning pace to integrate this distinct information with existing information already stored in the neocortex. Hence new declarative information is initially mainly represented in the hippocampal system with stronger representational 'weights' and in the neocortical system with weaker representation 'weights'. However, the above changes over time whereby the strength of representation 'weight' shifts from the hippocampal system to the neocortex (McClelland et al., 1995; O'Reilly & Norman, 2002). Unlike previous models of memory consolidation, the CLS makes direct predictions for new declarative information beyond that of memory consolidation on individual hippocampal-dependent episodic items to integration of the items with existing long-term neocortical knowledge.

Key evidence on how the CLS may potentially explain sleep-dependent consolidation and learning comes from behavioural studies mentioned in section 1.3.2 (see previous elaboration of Davis, Di Betta, Macdonald, & Gaskell, 2009; Dumay & Gaskell, 2007; Tamminen et al., 2010) and further neuroimaging studies of novel word learning (see Davis & Gaskell, 2009 for a more extensive review of this literature). An example of neural evidence for the CLS can be depicted by changes in brain activation to novel words across learning. In a recent study, experimenters found that after a night of sleep, novel words trained on the previous day elicited similar brain activation as existing words. In contrast, novel words that were not trained prior to sleep elicited significantly different responses compared to existing words (Davis et al., 2009). The general systems consolidation and CLS model attributes sleep-dependent consolidation and learning as a process occurring over multiple nights of sleep (McClelland et al., 1995). However, the behavioural and neural findings from the above studies have suggested that even a night of sleep can be beneficial towards the integration of new memories into existing knowledge. In addition schema learning research by Tse et al. (2007, 2011) has proposed that knowledge integration can occur rapidly and without sleep (see section 1.3.4).

In response to the contradictory findings regarding the time-course of memory consolidation in the CLS, McClelland (2013) proposed an adapted CLS model that incorporated 'rapid neocortical learning'. In the adapted CLS model, McClelland suggested that the time-course of memory consolidation depended on how consistent the newly learnt information was in relation to existing long-term information in the neocortex, where the more consistent the new knowledge to existing long-term knowledge, the quicker the transfer of information from the hippocampal to neocortical system. Based on evidence from multiple stimulations, McClelland concluded that the above findings suggesting rapid neocortical learning were aligned with the adapted CLS model that takes into account the level consistency of new knowledge with existing knowledge.

1.6.5 Synaptic homeostasis model

Many scientists also subscribe to the synaptic homeostasis model which hypothesizes that sleep helps normalize changes in synaptic strength across the brain that occurs during wakefulness, whereby memory consolidation in the synaptic homeostasis model is highly associated with slow-wave activity (SWA), Delta waves of < 4Hz. According to the model, the learning of information during wakefulness causes an increase in synaptic strength (synaptic potentiation). SWA during NREM sleep however leads to an effect of 'synaptic downscaling', where the downscaling of synaptic strength prunes away weak and noisy connections, and hence increasing the efficiency and strength of the remaining connections to potentially enable effective consolidation (Tononi & Cirelli, 2003).

Huber, Ghilardi, Massimini, and Tononi (2004) trained participants on a motor learning task with an embedded rotation adaptation component. Participants were then tested after they were trained and also in a second test session after 8 hours of sleep or wakefulness. Experimenters found that participants who had slept showed greater improvements in performance at the second test session compared to those who remained awake. Furthermore, the amount of improvement in the sleep group was positively related to the increase in SWA activity during sleep, giving further evidence that SWA is important for consolidation and learning.

Still, similar to many models, the synaptic homeostasis model has yet to give much insight on the role of sleep on knowledge integration. In addition, the model also predicts that weak associations will not be consolidated as well as associations that are strongly encoded. However, previous research on sleep and relational memory contradicts the above prediction as items that were not strongly associated during learning benefited more from sleep (Ellenbogen, Hu, Payne, Titone, & Walker, 2007; mentioned in section 1.3.2). Moreover, some studies from section 1.4 also indicated greater post-sleep benefits to weakly encoded items as compared to strong encoded items.

1.7 Summary and Thesis outline

The broad aim of this thesis is to investigate the effects of sleep on declarative memory consolidation and integration. Based on previous findings, this thesis aims to investigate whether particular components of sleep, such as sleep spindles and slow-wave sleep will play a direct role in benefitting knowledge integration. The thesis will focus on two main domains of declarative learning, namely word learning and numerical cognition. Although both aspects of learning are evolutionarily important, they have been shown to involve and activate different brain regions (Klimesch, Pfurtscheller, Mohl, & Schimke, 1990; Thioux, Pesenti, Costes, De Volder, & Seron, 2005), hence sleep may or may not have a different effect on each aspect of learning.

Based on current behavioural and neural evidence, the general two-stage systems consolidation model and CLS appear most parsimonious in providing insights on the effects of sleep on knowledge integration. Still, it is likely that the above models are not mutually exclusive and further research is needed to expand existing models to fully explain neural changes accounting for the effects of sleep in knowledge integration. This thesis also aims to provide insights and evidence as to which theoretical consolidation model best explains the effects on sleep on knowledge integration, if existing models need to be expanded or if a new scientific model needs to be developed to explain findings.

In addition, selective factors that modulate sleep-dependent consolidation, such as encoding strength and saliency of stimuli, have either not been widely researched in sleep literature or existing research has yielded conflicting results. Therefore, this thesis will examine if sleep-dependent consolidation of declarative memory would be modulated by between subject and within subject manipulations of encoding strength and saliency of learnt items, to better understand the implications of sleep on memory and learning.

Chapter 2: Overnight Sleep and Knowledge Integration in New Word Learning

2.1 Introduction

The term ‘vocabulary spurt’ is often used to describe the acceleration in the rate of word learning at around 18 to 24 months of age (Gopnik & Meltzoff, 1987; Ganger & Brent, 2004; McMurray, 2007). It is estimated that an educated adult has a vocabulary size of at least 60,000 words, most of which are acquired during childhood (Aitchison, 2003). Still, acquisition of new words is not uncommon in adulthood, as technological advancements have led to the development of new words (e.g. blog, cyber-bullying, hashtags). Adults also learn new words through the acquisition of new knowledge, for example students training to be doctors have to learn new medical terms. However, the most common instance of learning multiple new words during adulthood occurs in second language acquisition. Francis (1999) reviewed a series of experimental studies on bilingual language integration, which covered experimental studies using various methods including lexical decision and Stroop-like tasks. Francis (1999) suggested that the reviewed literature generally supported the notion of a shared first language and second language system for semantic representations. Therefore, a useful way to investigate the effects of sleep on knowledge integration in word learning might be through the learning of second language word meanings (semantic representations).

Two classic paradigms that are argued to demonstrate rapid and automatic access to word meanings are the size congruity effect and the semantic distance effect. The size congruity effect is similar to the classic Stroop effect (Stroop, 1935) and occurs when response times (RTs) are significantly faster for congruent compared to incongruent items that vary in referent sizes. A pioneering experiment on the size congruity effect was conducted by Paivio (1975), who presented participants with separate picture and word item-pairs differing in physical (size of pictures/height of words) and semantic (referent) size. Each item-pair was either congruent or incongruent. An item-pair was congruent when the semantically larger item was also physically larger (e.g. ZEBRA LAMP). In contrast, an item-pair was incongruent when the semantically smaller item was physically larger (e.g. LAMP ZEBRA). Participants were instructed to ignore the physical size of items and select the semantically larger item. Results revealed that for pictorial item-pairs, participants were significantly faster in judging congruent compared to incongruent pairs. In contrast, there was no effect of congruity for word-pairs. Even though results from the word-pairs did not mirror those of the picture-pairs, Paivio's findings were unlikely to be due to the notion that the word-pairs did not induce size congruity effects. Rather, the results might have been confounded by Paivio's experimental design where 54 word-pairs of different lengths were used and had repeated presentations. Subsequent research has indicated that differences in word length affect participants' RTs to the words (Dehaene, 1996; Moyer, 1973; Rubinsten & Henik, 2002), where in general, participants had swifter responses to words with shorter lengths. Therefore, Paivio's finding may have been confounded by his experimental design where individual words of different lengths were paired together.

Foltz, Poltrock, and Potts (1984) adapted Paivio's experiment where in one of Foltz et al's conditions participants were shown repetitions of six fixed word-pairs. The experimenters also added another condition where participants saw unique word-pairs (none of the words were repeated). Results from the fixed word-pairs condition replicated Paivio's finding where no effect of size congruity was present. However for the unique word-pairs, results revealed a large size congruity effect where congruent pairs were processed significantly faster than incongruent pairs. In addition, Foltz et al. also divided the 192 experimental trials with unique word-pairs into 4 blocks. Results revealed that even though there was a main effect of congruity for all blocks, the strength of the size congruity effect was much larger in the first compared to the last block. Based on the above finding, Foltz et al. highlighted that size congruity effects can occur in words and that Paivio's (1975) findings were influenced by his design of showing repetitions of fixed word-pairs. Participants in Foltz et al.'s study were not able to ignore instances where there was a disparity between the size of word-pairs and their semantic meaning (despite being instructed to ignore physical size). The above results suggest that semantic information is automatically processed when participants view written words.

In addition to the size congruity effect, when comparing items from the same category-dimension (e.g. animal names, object names, single digit numbers), the semantic distance effect often occurs, whereby RTs are faster for items that are further apart than for those that are closer on the relevant dimension (Van Opstal, Gevers, De Moor, & Verguts, 2008). The semantic distance effect was first documented by Moyer & Landauer's (1967) research which was based on number comparison, where participants were asked to select the numerically larger number from all possible pairs of single digit numbers. The experimenters found that participants' RT was negatively correlated with numerical distance, i.e. the RT for judging the larger number out of 1 and 9 is faster than the RT for 1 and 4. Evidence for semantic distance effects in words stems from Moyer (1973) who researched animal size comparison, whereby participants were presented with written animal names in pairs, e.g. 'ANT-COW' and had to judge which animal was semantically larger. In addition, participants were asked to estimate each animal's size on a numerical scale. Analysis of results showed that participants' RTs were significantly correlated with their approximation of animal sizes, such that the larger the distance between each animal pair (based on size approximation) the faster participants were at selecting the semantically bigger animal. The semantic distance effect has been replicated in many recent studies that include numeric and alphabetic stimuli (Van Opstal et al., 2008). The range of evidence supporting the semantic distance effect suggests that when comparing item-pairs within the same category-dimension, differences in semantics affect the speed at which the items are processed.

Rubinsten and Henik (2002) investigated both phenomena using Stroop-like tasks (Stroop, 1935) where known animal item-pairs differed in both their physical size (height of words) and semantic size (of the words' referents). Participants were asked to select the physically (font height) or semantically larger animal from item-pairs in separate comparison tasks. Results indicated that the semantic distance effect was displayed in semantic comparisons only, whereas the size congruity effect was displayed in both physical and semantic comparison tasks. The experimenters highlighted that the results from the above comparison tasks revealed that when processing words, meanings were rapidly and automatically activated, as participants' responses were affected by the lack of congruence between the font sizes of word-pairs and their meanings. It should also be noted that the differences in meanings affected processing of word-pairs only when the meaning was directly relevant to the task demands (in the semantic comparisons)[1].

[1] Rubinsten and Henik explained the lack of semantic distance effects in the physical comparisons by arguing that even though the relationships between form-meaning mappings of animal names were mentally represented in a continuum according to magnitude, they were done so in an abstract 'gist-like' manner. Hence, the disparity between meanings in word-pairs did not significantly contribute to differences in reaction times in the physical comparison tasks where they were not directly relevant to performance.

Tham, Lindsay and Gaskell (2011) adapted the paradigms from Rubinsten and Henik (2002) into a second language learning study where participants learnt six Mandarin characters referring to different-sized animals. The learning session was conducted either in the evening (wake group) or morning (sleep group) and participants were then tested after 12-hours of sleep or wake. During testing, both semantic and physical comparison tasks were conducted using the newly learned Mandarin characters. Equivalent tasks using English stimuli also provided baseline performance and controlled for circadian effects. The main finding was that the size congruity effects for semantic comparisons of the newly learnt Mandarin words were found for the sleep group but not the wake group. The main effect of congruity was present in all English stimuli and there were no significant differences in the strength of these effects between wake and sleep groups for the English stimuli. Hence, the newly learnt words displayed properties typical of established words only after participants slept but not when they were awake for a similar duration. Both the wake and sleep groups displayed significant semantic distance effects in both the English and Mandarin semantic comparisons. This suggests that sleep had a role in strengthening the semantic distance effect. For the participants who carried out the semantic judgement first, the sleep group showed a stronger distance effect than the wake group.[2] Still, even though results from Tham et al. (2011) provided initial support towards an association between sleep and knowledge integration for new word learning, the behavioural findings alone might not be sufficient as evidence towards an argument for an active role of sleep in knowledge integration.

In addition to behavioural data, sleep polysomnography data may be useful in providing evidence as to whether offline consolidation after learning serves to passively protect new information from interference or if sleep directly benefits memory consolidation of the new items. There has been growing evidence on the role of sleep spindles on declarative memory consolidation and knowledge integration. As mentioned in Chapter 1, Tamminen, Payne, Stickgold, Wamsley, and Gaskell (2010) reported that sleep spindle activity was positively related to the integration of newly learnt words with existing knowledge. This is because the experimenters found that the participants' overall spindle count was positively correlated with increased magnitude of lexical competition (an indication of the extent a novel word has been integrated to the mental lexicon) experienced by the novel words between testing sessions. The above finding is supported by previous

[2] However, in the absence of higher order interactions, this effect should be treated with caution.

research indicating the sleep spindle activity benefits declarative word-pair association tasks (Gais et al., 2002; Schabus et al., 2004), but is one of the first key studies reporting an integrative role for sleep spindles. Tamminen et al. (2010) also found that slow-wave sleep (SWS) was related to the strengthening of individual word memories as participants who experienced more SWS also showed greater improvements in speed of categorisation as to whether a word was a previously trained word or a foil. In summary, based on the above literature, it seems plausible to expect that sleep, especially specific components occurring during sleep such as SWS and sleep spindles, will lead to enhancements in the integration of new declarative information with existing knowledge.

The relationship between sleep and knowledge integration has been supported by many theoretical models. One such model is the complementary learning systems (CLS) model that was reviewed in Chapter 1. The CLS model draws from a general two-stage systems consolidation memory model (Davis et al., 2009; Davis & Gaskell, 2009; Diekelmann & Born, 2010a; McClelland et al., 1995), which incorporates hippocampal and neocortical memory systems. The hippocampal system is able to encode and learn information quickly, thus functioning as a temporary, initial store for distinct new memories. In contrast, the neocortical system acts as a slow-learning but long-term store that represents shared information across memories (Frankland & Bontempi, 2005; O'Reilly & Norman, 2002). The CLS proposes that new memories are encoded in both hippocampal and neocortical systems through interconnected networks, but rely more heavily on the hippocampal system connections during initial learning. However, when the brain is in an offline state such as sleep, the new memory representations are reactivated. The process of reactivation which occurs mainly during SWS leads to a reorganization of representations where increased reliance is placed on connections in the neocortical networks. Further offline reactivation leads to greater reinforcement of neocortical connections, where memory representations are further integrated into existing networks in the neocortical system and eventually become independent of the hippocampal system during retrieval (Diekelmann & Born, 2010a).

2.2 Experiment 1

Using size congruity and semantic distance effects as hallmarks of established and integrated word representations, Experiment 1 explored the relationship between overnight sleep and knowledge integration in new word learning. Based on the systems consolidation account, for integrated word meanings of newly learnt words, it is predicted that the new words should begin to exhibit typical size congruity and semantic distance effects. The new words in this experiment were learnt using a second language paradigm where the meaning of each word was explicitly taught, using Malay words referring to English animal names. The semantic comparison tasks used in Rubinsten and Henik's (2002) study were adapted to incorporate these newly learnt Malay words. The procedure of the study was similar to Tham et al. (2011) where participants learnt the new words in the evening (sleep group) or morning (wake group) and were tested approximately 12 hours later after a night of sleep or an equivalent duration of wakefulness. It was hypothesized that if sleep benefitted integration of new words, participants who slept between learning and testing would display greater size congruity effects and semantic distance effects for comparisons of the new Malay words than those who remained awake. In addition to testing with the new words, participants were also tested using established English words that were already fully consolidated in memory in order to control for circadian confounds and fatigue. If effects found with the Malay words were due to sleep-associated consolidation, then one should not expect to find differences between the two groups of participants for the English (baseline) stimuli.

A further aim of the experiment was to examine if specific components of sleep architecture (as reviewed in Chapter 1 and section 2.1) would lead to selective benefits for memory consolidation and knowledge integration. Participants in the sleep group remained in the sleep lab after training and polysomnography data was collected during sleep. This allowed the experimenter to examine whether specific components of sleep such as sleep spindles and (SWS) activity, that are key to the reactivation and transfer of memories from the hippocampal to neocortical networks in systems consolidation models, would also be related to selective benefits in size congruity and semantic distance effects. By incorporating sleep polysomnography, Experiment 1 aimed to expand on the behavioural findings in Tham et al., (2011) towards an argument for an active role of sleep in knowledge integration. Based on previous sleep and language research by Tamminen et al. (2010), the third hypothesis was that both sleep spindle and SWS activity would play a facilitative role in consolidation and integration of the new Malay words.

In addition to congruity and distance effects, Experiment 1 also explored measures of variations in participants' speed of response. In their work on second language word recognition, Segalowitz and Segalowitz (1993) developed a coefficient of variation in RT (CV_{RT}) by dividing a participant's standard deviation of RT by their mean RT. Segalowitz and Segalowitz (1993) proposed that the CV_{RT} was a marker of the level of automaticity in processing where a reduction in CV_{RT} would suggest an increase in automaticity, whereas an absolute reduction in RTs may not signal increased automaticity. Segalowitz and Segalowitz (1993) highlighted that a speed-up in absolute RTs would suggest that an individual was performing using the same mechanism but with more stability, hence leading to a quantitative benefit but no qualitative change in performance. In contrast, reduction in the CV_{RT} would signal that an individual was performing more rapidly due to a reduction or removal in previous 'controlled' processes. Therefore, a change in CV_{RT} denotes a change in the quality of performance due to increased automaticity with the reduction in previous 'controlled' processes.

In a related study on reading comprehension, Fukkink, Hulstijn and Simis (2005) trained Dutch secondary school students on an English language course, where 40 target English words were learnt during the course. The participants also completed two word recognition tasks, one before and one after training. Results revealed that participants had significantly smaller CV_{RTS} for the target English words at a post-training test compared to pre-training test session. The researchers suggested that the findings demonstrated participants' improved automaticity in processing the English words due to training that occurred between both sessions. To further investigate the role of sleep in integrating word meanings, Experiment 1 also aimed to explore differences in participants' of automaticity in processing using the CV_{RT} measure. Current findings suggest that well integrated items would also exhibit smaller CV_{RT}. However, unlike the existing research on CV_{RT}, it may not be reliable to compare CV_{RTs} during training and testing as different tasks were performed in each session. In contrast, it may be more reliable to compare CV_{RTs} for the new Malay word-pairs with the CV_{RTs} for the well-integrated English word-pairs. If sleep benefits integration of new word meanings, it is predicted that the sleep group would not exhibit any difference in CV_{RTs} for the Malay and English word-pairs whereas for the wake group, the CV_{RTs} for the Malay words would be significantly larger than the English words.

2.2.1 Methods

Apparatus and Materials

The test apparatus consisted of a computer with a 15-inch screen. The DMDX software (Forster & Forster, 2003) was used to collect all RT and error data. Participants used the 850F-Vibraforce USB-joypad for all experimental responses.

Experimental stimuli were chosen from 60 animal names from Paivio's (1975) size norm ratings. The 60 animal names were inputted into the English Lexicon Project database (Balota et al., 2007) where mean lexical decision RTs (M = 654 ms, SD = 85) were calculated. Word familiarity ratings (M = 480, SD = 62) were also calculated using the MRC Psycholinguistics database (Coltheart, 1981).Eighteen animal names were selected; all selected stimuli had lexical decision RTs and familiarity ratings within 2 standard deviations from the means of the 60 animal names. The selected words were either monosyllabic or disyllabic and were three to five letters long. The selected animal names were ranked according to mean size norms and allocated to 3 size groups: small (M = 1.41, SD = 0.45), medium (M = 3.81, SD = 0.66) and large (M = 6.63, SD = 0.55), with 6 animals in each group. The animal names were translated into Malay (alphabetic language) to provide the novel stimuli. All the Malay words were disyllabic and ranged from four to six letters per word (see Appendix 2 for the animal names and Malay translations).

Items in each size group were matched whereby an Analysis of Variance (ANOVA) revealed no significant differences in lexical decision RTs, familiarity ratings and word length (English and Malay stimuli) between animal-size groups (all p's > .05). Next, the items in each size group were randomly distributed into 2 sets, A and B. In summary, the combined stimuli consisted of 18 English and 18 corresponding Malay animal names (Appendix 2). Eighteen photographs measuring 45x30mm were selected from Google images to depict each animal in both sets. Photographs were matched whereby the animal took up approximately ¾ of the total area and all backgrounds were neutral (natural habitat).

A measure of general sleep quality was taken using the Pittsburgh Sleep Quality Index (PSQI;Buysse, Reynolds, Monk, Berman, & Kupfer, 1989), a questionnaire aimed to look at current sleep quality and also sleep quality for the previous month. The questionnaire included 7 components, namely subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep distances, usage of sleep medication and daytime dysfunction. A global PSQI score can be derived by adding the individual scores of the 7 components whereby clinical tests have indicated that a global score greater than 5 points to poor sleep quality. The Stanford Sleepiness Scale (SSS; Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973) was used to evaluate how alert participants felt both when they learnt and were tested on the new words. Participants had to indicate how sleepy they felt on a scale from 1 to 7, where 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.

Design & Stimuli

Experimental trials consisted of exposure, feedback and test trials (see Figure 1). There were 9 exposure trials for each stimulus set. In each exposure trial, a Malay word was centred on the upper half of the screen. Its corresponding animal name and photograph were located in the bottom left and right quadrants respectively (centre-to-centre distance: 70mm) and all words were displayed using Consolas font 90.

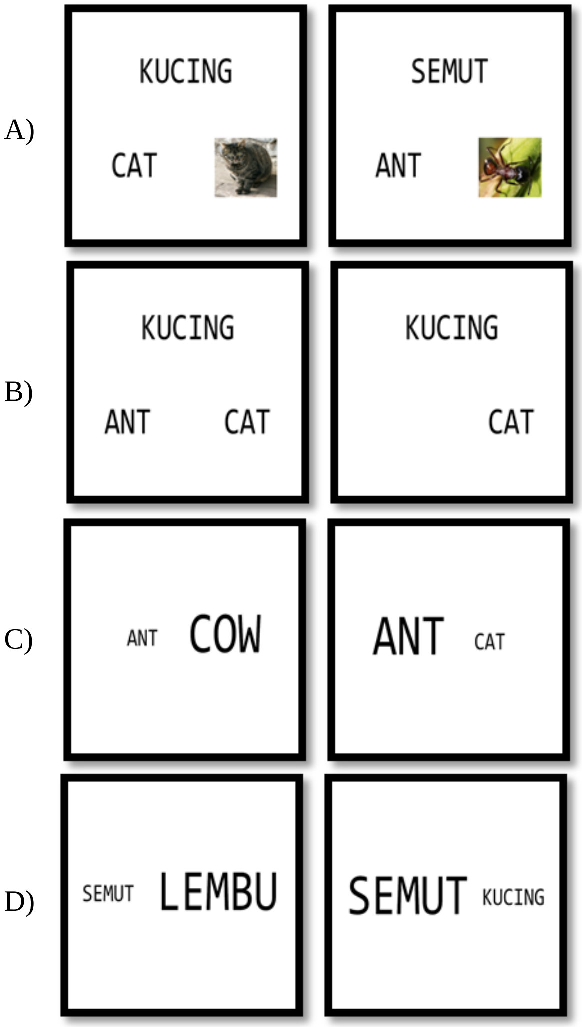


Figure 1. Experiment 1: Examples of experimental trials

A) Exposure, B) Feedback, C) Test (Control/English pairs, Congruent-large semantic distance vs. Incongruent-small semantic distance), D) Test (Malay word pairs, Congruent-large semantic distance vs. Incongruent-small semantic distance).

For the feedback trials, a target item was centred on the upper half of the screen and two items were displayed in the bottom left and right quadrants (centre-to-centre distance: 85mm). If the target was a Malay word, the two items in the quadrants were animal names, and if the target was an animal name, the two items were Malay words. Similarly to the exposure trials, all words were displayed using Consolas font 90. There were 6 feedback trials for each target animal name, where one of the two items always had the same meaning as the target (correct item) and the other was a foil. The 6 possible foils were made up of items from the same exposure set (2 small, 2 medium and 2 large animals). Each correct-foil combination was also presented twice, counterbalancing correct-foil positions (left/right). Items were also counterbalanced such that each animal name (English/Malay) appeared with the same frequency - 24 times (including correct-foil position counterbalancing); 12 times as the correct item and 12 times as foils. In summary, participants were exposed to 216 trials in total: 18 (9 English, 9 Malay) target animals x 6 feedback trials x 2 (left/right counterbalancing).

For the test trials, all possible pairs of stimuli were generated for the Malay and English animal names. Item-pairs were manipulated in a manner similar to that in Rubinsten and Henik's (2002) study, where items in each pair differed in physical font size and semantic size distance. For physical size difference, items differed in font height by either 1mm (one item was 6mm and the other 7mm) or 4mm (one item was 7mm and the other 11mm). For semantic size distance, items differed by either small (small-medium, medium-large animal) or large (small-large animal) distances. The centre-to-centre distance between words was 50mm.There were 18 possible word-pairs differing by small semantic size distances and 9 item-pairs differing by large semantic size distances for each stimulus set presented in the same language (English/Malay). Each small and large semantic size distance-pair was also manipulated in terms of physical size difference. Each possible pair was then presented twice, counterbalancing items on the left and right to prevent bias in response. To ensure that participants viewed the same number of small semantic and large semantic size distance pairs, each counterbalanced large semantic size distance pair was then presented twice. Congruity was manipulated whereby half the trials were congruent and half were incongruent. Therefore, for each comparison task, 288 trials were presented in a random order for each participant: 36 (small and large semantic size distance-pairs) x 2 (physical size difference manipulations) x 2 (left/right counterbalancing) x 2 (congruity); giving a total of 576 trials for both English and Malay comparison tasks combined.

A six-way mixed-subjects design was used. The dependent variable was the RTs for correct responses. There were six independent variables (IV): two between-subjects and four within-subjects. The between-subjects IVs were experimental group (wake/sleep) and stimulus set (A/B). The within-subjects IVs were type of language (English/Malay), congruity (congruent/incongruent), physical size difference (1mm/4mm) and semantic size distance (small/large). A separate six-way ANOVA was conducted with participants' error rates as a dependent variable and presented in Appendix 1.1.

Participants

Thirty two monolingual native English speakers (13 males, 19 females; *M* = 21, range:18-34) who were non-smokers with no known language disorders and no prior history of drug or alcohol abuse, neurological, psychiatric or sleep disorders (screened by the PSQI) were recruited via an online database and were given a cash reward or course credit for their participation. Participants were randomly allocated to the awake (n = 16, 9 females) or sleep (n = 16, 10 females) experimental groups. All participants had a sleep-wake pattern where they would typically rise by 9am each morning and were used to having at least 6 hours of sleep per night or longer. Participants were also emailed an information sheet before the experiment instructing them to have a night of normal sleep (based on their usual sleeping patterns) prior to the day of the experiment. In addition, they were required not to consume caffeine during the experiment and to refrain from taking recreational drugs or alcohol 24 hours prior to the training session and also during the experiment. Participants also agreed to wake by 9am on the day of testing. The study was approved by the Department of Psychology (University of York) ethics committee.

Procedure

Each participant underwent a training and a test session, both carried out in the University of York Psychology departmental sleep laboratory. The procedure within both sessions was identical in sleep and wake groups with training-test interval of approximately 9 hours. At the start of the training session, participants filled in a consent form, the PSQI and SSS. They also filled in a SSS at the start of the test session. Throughout the sessions, participants were seated at a viewing distance of approximately 60cm from the screen. The main difference between experimental groups was the timing of the sessions, which will be further discussed below. A summary of the experimental procedure can be found in Figure 2.

Overview of Procedure

Training Session (9 new Malay words)
1) Exposure phase • 3 exposures per word 2) 2-AFC Feedback Phase • 24 exposures per word, 12 as targets and 12 as foils

Test Session
1) English Comparisons (baseline circadian control) • 288 trials 2) Malay Comparisons • 288 trials with word-pairs consisting of Malay words learnt during the training session divided into 144 congruent and 144 incongruent trials • 18 unique small semantic distance word-pairs • 9 unique large semantic distance word-pairs (repeated twice)

Timing of Sessions for Wake (white columns) and Sleep (grey columns) groups

Training Session	Daily Activities (excluding sleep)	Test Session					
0930		1930	2100	2200	2300	0715	0800
			PSG Setup	Training Session	Sleep (Lights off time)	Wake	Test Session

Figure 2. Summary of Experiment 1 procedure including timings of sessions and tasks carried out in the training and test sessions.

Participants in the wake group arrived in the laboratory at approximately 9:30am. They then underwent a training session that began at about 9:40am. Participants returned for the test session between 7:30-7:45pm on the evening of the same day. Between sessions, participants were allowed to carry on with their daily activities, with the exception of taking naps and consuming caffeinated or alcoholic products. For the sleep group, participants arrived at the laboratory at approximately 9 pm. Prior to the training session, participants underwent an electrode setup for the overnight polysomnographic recording. The training session then began at about 10pm and lasted for approximately 20 minutes. After that, participants remained in the sleep laboratory for a minimum period of 8 hours, which included time spent in overnight sleep. To prevent confounds from effects of sleep inertia, the test session took place at least 30 minutes after participants awoke. In addition, at the end of the test session the vocabulary and matrix reasoning subtests from the Wechsler Abbreviated Scales of Intelligence (WASI) (Wechsler, 1999) were administered.

Training Session. For both groups, half the participants were randomly selected to be trained using Set A stimuli pairings and half using Set B pairings. Participants were first shown exposure trials. There were 9 different trials representing the 9 Malay words in the stimuli set. These 9 trials were repeated in 3 separate blocks and presented in a random order within each block. Each trial was displayed for 5000ms and was followed by a subsequent trial after a 1000ms inter-trial-interval (blank screen), where no response was required. Next, participants were shown 2AFC feedback trials consisting of a target and two items. Of the two items, participants had to select the item with the same meaning as the target. Trials were presented in a random order and displayed for 4000ms or until participants responded. 500 ms after responding, participants were given feedback whereby the incorrect item 'disappeared', leaving the target and correct item on the screen. Feedback was given regardless of whether the response was correct or incorrect. Participants were given the opportunity to take an untimed break half way through the task.

Test Session. Participants carried out two comparison tasks in a test session: one with known English stimuli to control for circadian effects and one with the Malay words. The English task consisted of animal names from the opposing stimuli set and the Malay task consisted of words that participants learnt in the training session. That is, if participants learnt stimuli pairings from Set A, they viewed Malay trials with words from Set A and English trials consisting of animal names from Set B and vice versa. Participants were instructed to select the semantically 'larger' item from item-pairs, as quickly and as accurately as possible. 'Larger' always referred to the semantic size of the animal the item was depicting, regardless of the font size of the item. Participants always completed the English task first, followed by the Malay task. For the tasks in English, participants were given four practice trials. There were no practice trials for tasks in Malay. Before each trial, a fixation cross '+' was displayed in the centre of the screen for 500 ms. Each trial was then presented for 4000ms or until a response was made. Participants were given the opportunity to take an untimed break half way through and the end of each task.

Polysomnographic (PSG) recording. Polysomnographic data were recorded using an Embla N7000 system and Remlogic software. Electrodes were set up following the International 10-20 system (as recommended by Rechtschaffen & Kales, 1968). A 10 channel montage was used, with four scalp electrodes referenced to contra-lateral mastoids (C3-M2, C4-M1, O1-M2 and O2-M1), left and right electro-oculographic (EOG) channels and 2 chin electromyographic (EMG) channels. All technical and digital specifications, including impedance levels, sampling rates and filter settings were set according to the recommended specifications in the standardised American Academy of Sleep Science Manual (AASM; Iber, Ancoli-Israel, Quan, 2007). Data were scored manually in 30-second epochs according to the Rechtschaffen and Kales (1968) sleep staging criteria, with one revised criterion - sleep stages 3 and 4 were scored collectively as SWS. Twenty percent of the data were cross-scored by a second independent scorer and the Remlogic generated inter-scorer reliability report revealed an average of approximately 80% overall score agreement between the 2 independent scorers.

Spindle analysis was carried out by taking the mean activity from the bilateral central electrodes (C3 and C4) using the EEGLab toolbox (Delorme & Makeig, 2004) in Matlab. Only NREM stage 2 sleep and SWS were included in the sleep spindle analysis as the majority of sleep spindles occur in these stages. Data from stage 2 sleep and SWS were analysed together as NREM sleep except in cases of statistical interest. Firstly, epochs containing any movements, arousal or noise artefacts were removed from analysis. Before extracting the sleep spindles, the EEG signal was band-pass filtered using a finite impulse response (FIR) filter such that only data between 11 and 15 Hz remained. Sleep spindles were then automatically detected by an algorithm adapted from Ferrarelli et al., (2007).

2.2.2 Results

Data from one participant in the sleep group were excluded due to disturbed sleep and awakenings for more than 50% of the polysomnography recording. There were no significant differences in SSS ratings between training and test sessions for the sleep and wake group (all p's > .27).

Throughout this thesis, all marginally significant effects are reported and are consistently interpreted as potentially meaningful.

Training Session

The mean error rate for feedback trials was 3.8%. Participants' error rates were arcsine square-root transformed (Bromiley & Thacker, 2002) and an ANOVA was performed on the transformed error rates with stimulus set and group as between-participant factors. This was done to check for learning-related confounds that may have arisen due to differences in the stimulus sets or time of training. There were no significant differences in error rates based on stimulus set [$F(1, 27) = .86$, $p = .36$] or time of training (i.e. wake or sleep group) [$F(1, 27) = .005$, $p = .94$].

Test Session

Table 1. Experiment 1: Mean response times for all correct responses after trimming. Standard errors are presented in parentheses.

	English		Malay	
	Wake	Sleep	Wake	Sleep
Congruity				
Congruent	904 (45)	918 (50)	1338 (82)	1157 (72)
Incongruent	912 (49)	946 (45)	1360 (82)	1187 (73)
Semantic Distance				
Small	953 (47)	983 (48)	1403 (85)	1237 (76)
Large	863 (47)	881 (47)	1294 (79)	1107 (69)

In order to detect outliers, RTs for congruent and incongruent trials for the English and Malay comparisons were separated (i.e. each participant had four sets of mean RTs). The RTs were not separated for all of the variable factors as too many data points would have been removed. The RTs were divided into language as there was a large main effect of language [$F(1, 27) = 70.34$, $p < .001$] due to substantial differences in RTs between the English and Malay trials. Next, the experimenter decided to focus on the congruity measure as the size congruity effect is thought to require stronger stimulus-response mapping and hence may theoretically be a purer measure of automaticity (see discussion in section 2.3).

For every set, scores more than 2 standard deviations away from each participant’s mean RTs for the corresponding task were discarded. This resulted in an average data loss of 1.7% and 1.4% per participant for the Malay and English comparisons respectively. The above method of detecting and trimming outliers was adapted from other language experiments (such as Balota et al., 2007; Gollan & Ferreira, 2009; Sereno, O’Donnell, & Sereno, 2009; Yap, Liow, Jalil, & Faizal, 2010). Mean RTs for all remaining correct responses were calculated and are presented in Table 1 (a comprehensive summary and breakdown of RTs from all factors can be found in Appendix 3).

Table 2. Experiment 1: ANOVA results of main effects and interactions with group and congruity or distance effects in semantic comparison tasks

Factors	**F**	**p**
Language (Lang)	70.27***	<.001
Congruity (Cong)	7.49*	.011
Semantic Distance (Dist)	233.33***	<.001
Physical Font Difference (Phy)	.51	.48
Group	.71	.41
Cong x Group x Lang	.075	.78
English (Cong x Group)	.83	.36
Malay (Cong x Group)	.11	.74
Dist x Group x Lang	.29	.59
English (Dist x Group)	.51	.49
Malay (Dist x Group)	1.04	.31
Cong x Dist x Group x Lang	5.36*	.028
English (Cong x Dist x Group)	.27	.60
Malay (Cong x Dist x Group)	5.93*	.025
Congruent (Dist x Group)	5.75*	.024
Incongruent (Dist x Group)	.19	.66

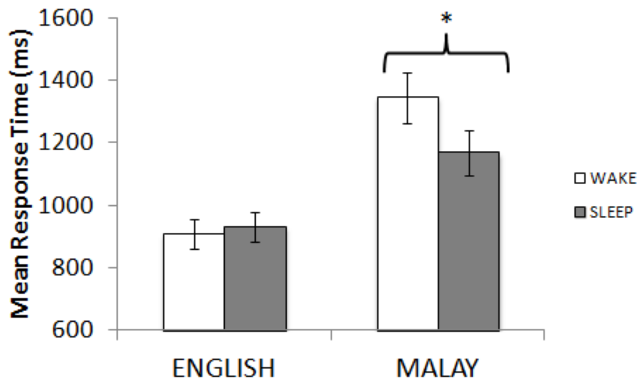
Cong x Dist x Phy x Group x Lang)	3.10#	.090
English (Cong x Dist x Phy x Group)	.34	.56
Malay (Cong x Dist x Phy x Group)	3.78#	.062
Small Dist	.47	.50
Small Phy (Cong x Group)	1.03	.32
Large Phy (Cong x Group)	2.04	.16
Large Dist	4.83*	.037
Small Phy (Cong x Group)	.01	.92
Large Phy (Cong x Group)	5.61*	.025

Note. *p < .05. ***p < .001, #p <.1

A six-way mixed ANOVA was performed with group and stimuli set as between-participant factors, and language, congruity, physical size difference and semantic size distance as within-participant factors. All main effects and interactions that included group and size congruity effects or group and semantic distance effects are displayed in Table 2. Main effects and interactions with the stimuli set were not reported (Pollatsek & Well, 1995). Results revealed a significant main effect of language whereby participants had faster RTs in the English compared to the Malay comparison task [$F(1, 27) = 70.34$, $p < .001$]. There was a significant interaction between language and group [$F(1, 27) = 5.85$, $p = .022$] whereby the sleep group had faster RTs than the wake group for the new Malay words but not for the English words (see Figure 3a).

To evaluate whether the above interaction reflected changes in knowledge integration of the new words rather than an absolute 'speed-up' of responses in the sleep group, a three-way mixed ANOVA was conducted on a measure of the CV_{RT} with language as a within-participant factor and group and stimuli set as between-participant factors. Similar to the RT analysis, there was a significant interaction between language and group [$F(1, 27) = 10.82$, $p = .003$] for the CV_{RT} analysis. Hence, it is unlikely that the interaction between language and group was only due to an absolute 'speed-up' of responses as participants in the sleep group had faster RTs and also smaller variations in their responses compared to the wake group for the new Malay stimuli (see Figure 3b). The CV_{RT} results also revealed that there were no group differences for the English stimuli. In addition, when looking at the wake and sleep groups separately, the wake group had significantly less variation in RTs for the English compared to the Malay comparisons [$F(1, 14) = 27.84$, $p < .001$]. In contrast, there was only a marginal significant difference between the variation in RTs in the English and Malay comparisons for the sleep group [$F(1, 13) = 4.20$, $p = .061$]. Therefore after a night of sleep, the variation in performance for the new words was highly similar to the variation for existing integrated words to large extent whereas for the wake group variation in performance for new words was significantly different to that to existing words.

(a)



(b)

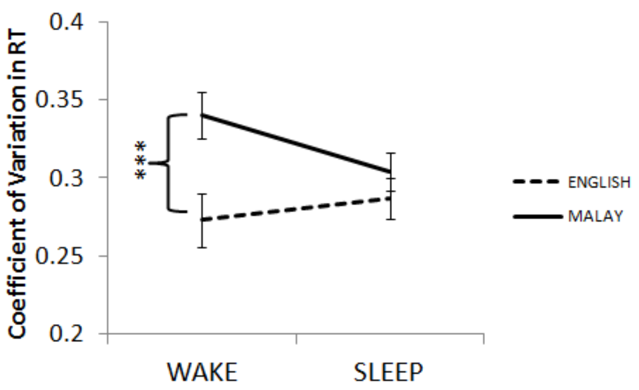


Figure 3. Experiment 1: (a) Mean response times for the English and Malay comparison tasks. (b) Mean CV_{RTs} for the wake and sleep groups for English and Malay words. Error bars represent standard error of the means.

Still, in order to further ensure that the CV_{RT} effect was not mainly driven by the differences in RT, a three-way mixed ANOVA was also conducted on participants standard deviations in RT (SDs) with the same factors as the RT analysis. There was a significant interaction between language and group [$F(1, 27) = 14.44$, $p < .001$] for the analysis on participant's SDs. There was no differences in SDs between the wake and sleep groups when looking at the English comparisons [$F(1, 27) = .77$, $p = .39$]. However, the sleep group displayed significantly smaller SDs than the wake group for the Malay comparisons [$F(1, 27) = 5.03$, $p = .030$]. This highlights that even though that were significant improved in RTs due to practise effects, the CV_{RT} results were also driven by group differences in RT variance measured by the SDs.

As there was a large main effect of language in the six-way ANOVA, in the cases where main effects and interactions were significant for the combined (English and Malay) ANOVA, further analysis was conducted to investigate whether the main effect and interactions were present in both languages or only in one language. This was to ensure that the results were not affected by the substantial difference in absolute RTs between languages.

The semantic distance effect was significant across the wake and sleep groups for both the Malay [$F(1, 27) = 101.13$, $p < .001$] and English [$F(1, 27) = 193.23$, $p < .001$] comparisons with RTs to item-pairs with large semantic distances faster than those item-pairs with small distances. The size congruity effect, where RTs for congruent trials were quicker than incongruent trials, was marginally significant for both the Malay [$F(1, 27) = 3.93$, $p = .058$] and English [$F(1, 27) = 2.93$, $p = .098$] comparisons[3].

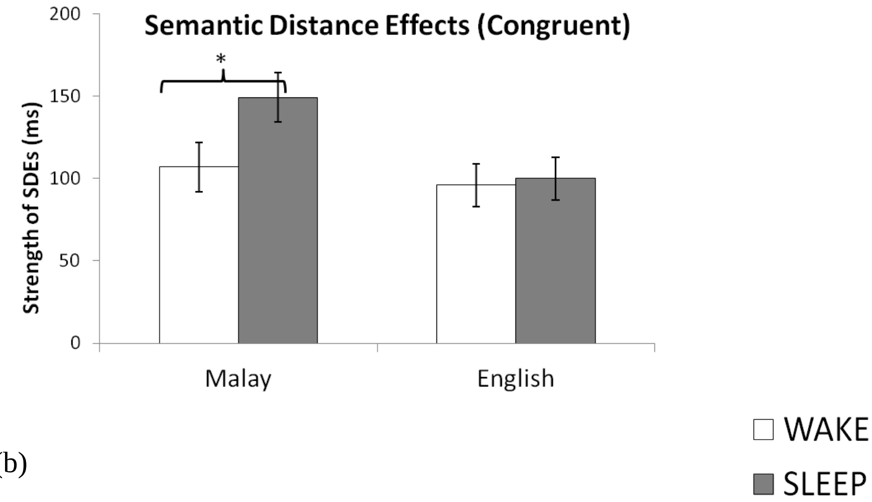
Importantly, the analysis revealed an interaction between congruity, semantic distance, group and language, [$F(1, 27) = 5.36$, $p = .028$]. Breaking this down, the interaction between congruity, semantic distance and group was significant in the Malay [$F(1, 27) = 5.93$, $p = .025$] but not the English [$F(1, 27) = .27$, $p = .60$] comparisons. This suggests that group differences were more pronounced for the new Malay words than the existing English words. When looking specifically at the results for congruent trials in Figure 4a, there was a

[3] Unlike the previous literature on size congruity effects, the English items were not matched for word length, hence the lack of size congruity effects in English comparisons in Experiment 1 (Dehaene, 1996; Moyer, 1973; Rubinsten & Henik, 2002).

significant interaction between semantic distance and group where the sleep group displayed a stronger semantic distance effect than the wake group for the Malay comparisons [F (1, 27) = 5.75, p = .024]. There were no equivalent interactions between distance and group for the incongruent trials [F (1, 27) = .19, p = .66].

In addition, there was a marginally significant interaction between congruity, physical difference, semantic distance, group and language [F(1, 27) = 3.10, p = .09]. When looking at the languages separately, the interaction between congruity, physical difference, semantic distance and group was marginally significant in the Malay [F(1, 27) = 3.78, p = .062] but not the English [F(1, 27) = .34, p = .56] comparisons, which suggests that the group differences are unique to the newly learnt Malay words. When considering all possible combinations of word-pairs, further analysis highlighted that the above interaction was driven by word-pairs with large semantic distances and large physical font size differences (p values were >.16 for the other combinations of word-pairs). For word-pairs with large semantic size distances and physical font size differences (4mm), there was a significant interaction between congruity and group where the size congruity effect was greater in the sleep compared to the wake group for the Malay comparisons [F(1, 27) = 5.61, p = .025]. Moreover, when looking at the wake and sleep groups separately, the sleep group displayed a significant size congruity effect [F(1, 13) = 7.50, p = .017], but the wake group did not [F(1, 14) = .27, p = .061] (see Figure 4b).

(a)



(b)

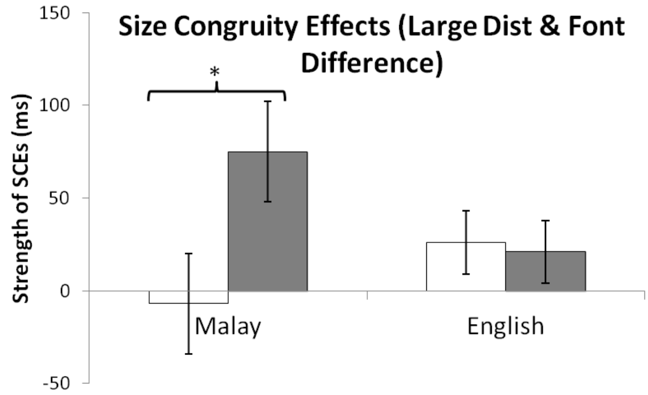


Figure 4. Experiment 4: Strength of automaticity effects: (a) Semantic Distance Effects, (b) Size Congruity Effects. Error bars represent standard error of the means.

All other main effects and interactions were not significant for both the Malay and English comparison tasks except the interaction between congruity, physical difference, and semantic distance for the English comparison [$F(1, 27) = 4.56$, $p = .042$]. The above finding was not elaborated as there were no significant interactions when comparing individual variables.

Sleep stage and Spindle analysis

Table 3. Experiment 1: Sleep parameters for participants in the sleep group. Parentheses denote the standard deviation of the means.

Sleep Parameter	Mean time in minutes	Time as a % of total sleep time
Total sleep time	486 (33)	
Wake time after sleep onset	18 (22)	
Sleep latency	25 (14)	
Stage 1	46 (21)	9.4 (4.1)
Stage 2	252 (22)	51.9 (5.0)
SWS	99 (26)	20.2 (4.8)
REM	90 (16)	18.5 (3.1)

Note: SWS= slow-wave sleep, REM= Rapid eye movement sleep

Table 4. Experiment 1: Mean sleep spindle measures for central electrodes. Parentheses denote the standard deviation of the means.

	C3	C4	Mean Central (C3 + C4)
Spindle Density	.061 (0.33)	0.51 (0.25)	0.57 (0.29)
Total Count	216 (88)	181 (88)	197 (100)

Note: Total count = total count of spindles, Spindle Density = mean spindle count per minute

In addition to analysis on the behavioural data, further analysis was also conducted on the sleep polysomnography data. The main sleep parameters of the participants in the sleep group are displayed in Table 3. The most common measures of reflecting spindle activity are depicted in Table 4, which is made up of the count of all NREM (stage 2 and SWS) spindles detected and the spindle density (mean number of spindles per minute in stage 2 sleep and SWS combined). Spindle activity for the C3 was discarded for one participant due to a fault with the electrode and for that participant only activity from the C4 electrode was included.

Table 5. Experiment 1: Correlations between measures of automaticity (sensitive to group differences) and time spent in different sleep stages (as a percentage of total sleep time) and NREM spindle count.

Language	Test Measure		Stage 2	SWS	REM	Spindle Count
Malay Comparisons						
	SCE (all trials)	*r*	0.10	-0.13	-0.006	**0.61**
		p	0.72	0.64	0.98	**0.015†**
	SCE (large physical and semantic size difference)	*r*	-0.21	0.005	0.05	**0.71**
		p	0.45	0.99	0.84	**0.003**
	SDE (all trials)	*r*	-0.42	0.51	-0.14	0.13
		p	0.11	0.05	0.62	0.63
	SDE (congruent trials)	*r*	**-0.63**	**0.66**	-0.39	0.35
		p	**0.013†**	**0.007**	0.15	0.19
English Comparisons						
	SCE (all trials)	*r*	0.34	-0.29	0.45	0.04
		p	0.21	0.29	0.08	0.86
	SCE (large physical and semantic size difference)	*r*	0.17	-0.23	0.45	0.19
		p	0.54	0.40	0.08	0.49
	SDE (all trials)	*r*	0.074	-0.14	0.43	-0.19
		p	0.79	0.60	0.10	0.48
	SDE (congruent trials)	*r*	0.034	-0.053	0.26	-0.005
		p	0.91	0.85	0.35	0.98

Note: Significant correlations in bold. † = *p*-values that do not survive a Bonferroni correction for multiple comparisons. SCE = strength of the size congruity effect, SDE = strength of the semantic distance effect

Measures that were sensitive to group differences were correlated with the percentage of total sleep time[4] spent in stage 2 sleep, SWS and REM sleep, plus NREM spindle count[5] (see Table 5). All correlations between behavioural results and sleep architecture underwent Bonferroni correction for multiple comparisons. The comparisons with test measures within each language task (see the Test Measure column in Table 5) were calculated as multiple comparisons. This resulted in a threshold p-value of .0125 after Bonferroni corrections. This method of correcting for multiple comparisons was adopted from the analyses made in Tamminen et al. (2010), where multiple corrections were based on contrast comparisons made in each task.

First, when considering the overall size congruity effects, there was a significant correlation between NREM spindle count and the strength of the size congruity effect across Malay comparisons, r = .61, p = .015, but this did not survive Bonferroni correction. Still it is important to note that there was no equivalent relationship between spindle count and size congruity effect for the overall English comparisons, r = .04, p = .84. Next, there was a significant Bonferroni-corrected positive correlation between spindle count and the strength of the size congruity effect in trials with a larger physical size difference and semantic size distance for Malay comparisons, r = .71, p = .003 but not for English comparisons, r = .19, p = .49, where the difference in the Steiger's Z score for both correlation coefficients approached significance, z = 1.91, p = .089 (see Figure 5a).

Analysis of semantic distance effects across all Malay trials with sleep stage and NREM spindle count did not reveal any significant relationships. However, in the more focused analysis of semantic distance effect for congruent trials (which showed effects of group for the Malay trials), there was a relationship between strength of semantic distance effect and the time spent in SWS, r = .66, p = .007. In contrast, there was no significant relationship for the equivalent comparison when looking at English congruent trials, r = -.053, p = .85 (see Figure 5b). In addition, Steiger's Z scores revealed a significant difference between the Malay and English

[4] It should be noted that that when looking at the absolute time spent in stage 2, SWS and REM sleep rather than as a percentage of total sleep time, there was no change in the pattern of findings.

[5] Correlations with spindle density were not included as they were strongly correlated with spindle count for stage 2 (r = .921, p < .001), SWS (r = .937, p < .001) and NREM (stage 2 and SWS combined) sleep (r = .990, p < .001), hence suggesting that both total count and density may reflect the same measure.

correlation coefficients, $z = 2.07$, $p = .038$. There was a significant negative correlation between the strength of the semantic distance effect for Malay congruent trials and the time spent in stage 2 sleep, $r = -.63$, $p = .013$, but the correlation did not survive Bonferroni correction. After further analysis, it seems likely that the above correlation with stage 2 sleep was driven by a general negative correlation between time spent in stage 2 sleep and SWS, $r = -.63$, $p < .001$.

In addition, partial correlations were conducted controlling for spindle count in the SWS analysis and Malay semantic distance effect (congruent) relationship, $r = .72$, $p = .003$, and controlling for SWS in the correlation between spindle count and Malay size congruity effect (large physical difference and semantic distance), $r = .71$, $p = .004$. The partial correlations revealed that SWS and NREM spindle count were uniquely related to semantic distance effect and size congruity effect respectively.

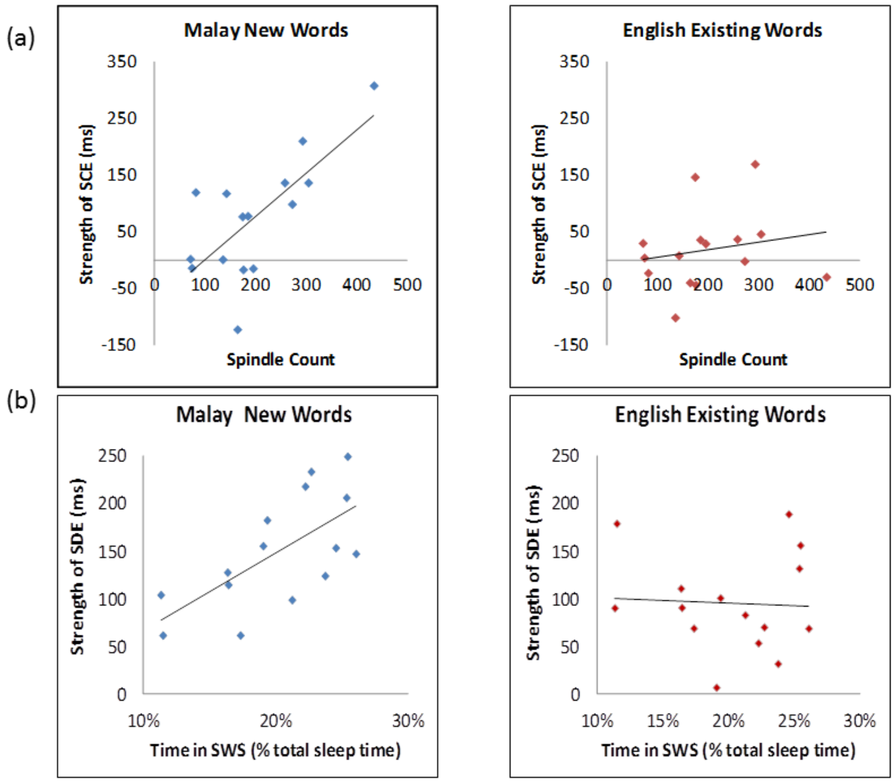


Figure 5. Experiment 1: Correlation between (a) Size Congruity Effects (Large Distance and Physical Difference) and NREM spindle count, (b) Semantic Distance Effects (Congruent) and SWS.

As previous research has highlighted that spindle activity was positively related to intelligence (Fogel & Smith, 2006, 2011), additional analyses were conducted by correlating standardized WASI Vocabulary, Matrix Reasoning and combined Vocabulary and Reasoning scores with sleep stage and sleep spindle activity. Bonferroni correction for multiple comparisons was carried out using the same method as for the correlations with behavioural findings. The only correlation which survived the Bonferroni correction was the positive relationship between SWS and the Reasoning sub-score, $r = .61$, $p = .015$. There was also a positive relationship between Stage 2 spindle count and Vocabulary, $r = .54$, $p = .037$, that did not survive Bonferroni correction. A partial correlation controlling for Reasoning scores revealed that the significant correlation between SWS and the Malay semantic distance effect (congruent trials) remained, $r = .70$, $p = .005$. Furthermore, the significant relationship between NREM spindle count and the Malay size congruity effect (large physical difference and semantic distance) also remained when controlling for Vocabulary scores, $r = .65$, $p = .012$. Both partial correlations indicated that the current experimental findings on sleep-associated knowledge integration were not confounded by regarding potential relationships between individual differences in intelligence and sleep architecture.

2.3 Chapter Summary and Discussion

Experiment 1 investigated the relationship between sleep and knowledge integration in newly learnt second language words, using size congruity and semantic distance effects. The main finding was that the sleep group displayed larger size congruity effects than the wake group for the newly learnt Malay word-pairs with large physical size differences and semantic size distances. When the wake and sleep groups were analysed separately, there was a significant size congruity effect in the sleep group but not in the wake group for those items. Hence, a single night of sleep (as compared to wakefulness) appeared sufficient to modify processing of meanings in a certain subset of newly learnt second language words, making them exhibit properties similar to existing English words. Results also highlighted that when making judgements of the newly learnt Malay words, participants in the sleep group experienced significantly stronger semantic distance effects than those in the wake group for congruent trials. This indicates that after just one night of consolidation, participants in the sleep group were affected by differences in semantic meanings when processing newly learnt words and that sleep enabled more rapid activation of semantic information in the new words. In general, participants who slept after learning showed behavioural effects for newly learnt words that were more typical of integrated English words than the effects displayed by those who remained awake. Moreover, there was no difference in performance accuracy in the training session 2AFC task between both the wake and sleep groups. Hence, the results were unlikely to be confounded by circadian effects such as time of training or testing, but rather reflect the notion that sleep benefits the processing of semantic meanings of newly learnt words but not of existing words that are already consolidated.

Another explanation for the above behavioural differences in performance between the wake and sleep groups for the newly learnt Malay words, but not for English words, could be that recently learnt memories were retrieved in a different manner from well consolidated memories. Therefore, instead of a sleep-associated effect, the stronger size congruity and semantic distance effects experienced by the sleep group for the Malay words may be due to a retrieval associated effect linked to recently acquired memories. Even though Experiment 1 did not directly separate sleep-associated from retrieval-associated effects, by incorporating sleep polysomnography data it was able to examine whether specific sleep architecture had a direct relationship with the behavioural size congruity and semantic distance effects for Malay words. This would better align findings with an argument for an active role of sleep in knowledge integration rather than a retrieval-associated effect. When comparing behavioural findings with sleep polysomnography data, the strength of the semantic distance effect for Malay congruent trials was positively correlated with time spent in SWS; while the size congruity effect was positively correlated with NREM spindle count for Malay word-pairs with larger physical size differences and semantic size distances. These findings suggest that sleep is not only associated with behavioural changes in processing new word meanings, but that some individual components of sleep may also have an active role in integrating these new form-meaning associations into existing knowledge.

It should be noted that differences between the wake and sleep groups in the size congruity and semantic distance effects did not occur for all Malay trials but were displayed only in trials with large physical differences and semantic distances (size congruity effect), or congruent (semantic distance effect) trials. Interactions between congruity and semantic distance have also been displayed in other studies. In their study on numerical cognition, Schwarz and Heinze (1998) found ‘facilitative’ effects of congruity on semantic (numerical) distance when participants were comparing established numbers, where participants showed larger semantic distance effects for congruent compared to incongruent trials. Since this relationship exists for established numbers, it seems plausible that sleep-associated benefits in semantic distance for newly learnt words would first emerge in congruent trials if participants only have one consolidation night. When considering the effects of physical font differences and semantic distances on size congruity, Cohen Kadosh and Henik (2006) found an additive effect of physical stimuli properties on size congruity in their study, where the strength of the size congruity effect increased with increased luminance difference between stimuli. This finding gives an explanation of the sleep-dependent benefit in size congruity effect for trials with larger physical differences. However, some studies on numerical cognition have found that the size congruity effect decreases with larger semantic (numerical) distance between items (Cohen Kadosh & Henik, 2006; Tzelgov, Meyer, & Henik, 1992). Experiment 1 has also shown the opposite trend of a sleep-dependent size congruity effect for newly learnt items with large semantic distances. One possible explanation for this difference could be that semantic distance is different in numerical cognition and word learning (there were no equivalent differences found in Rubinsten and Henik's (2002) study using word stimuli). It is also important to note that the above examples consist of stimuli with established (well-learnt) items, whereas the items learnt in Experiment 1 only underwent one night of consolidation. Hence, it is possible that after a longer consolidation period, the size congruity effect in Malay items may decrease with greater semantic distance (Cohen Kadosh & Henik, 2006; Tzelgov et al., 1992) or be independent of manipulations of semantic distance (Rubinsten & Henik, 2002).

Another explanation for the lack of effects across all Malay trials could be related to the fact that RTs are affected by levels of saliency. As mentioned in Chapter 1, saliency can be broadly defined as a factor that causes increased cognitive awareness such that the item or thought stands out (Myers & Alpert, 1977). Canli et al. (2004) found that both depressed and healthy participants were quicker at responding to emotionally salient words than neutral words. Even though the emotional saliency was not manipulated in the current experiment, congruent trials and trials containing word-pairs with larger physical size differences and semantic size distances were more salient than the other experimental trials based on Myers and Alpert’s (1977) definition, and may have led to unaccounted variability in participants RTs, especially for new Malay words that were not fully consolidated. Hence, additional testing after multiple nights of sleep would be useful to see if effects would emerge.

In addition to the main experimental findings on the size congruity and semantic distance effects, sleep was also related to overall swifter speed of judgments for Malay words, as the sleep group had faster RTs than the wake group across all Malay stimuli. There were no significant differences in overall RTs between the wake and sleep groups for the English stimuli, indicating that the beneficial effect of sleep towards comparisons of new words was not due to confounds such as circadian effects. Moreover, as mentioned in section 2.2, Segalowitz and Segalowitz (1993) proposed that the CV_{RT} was a marker of the level of automaticity in processing where a reduction in CV_{RT} would suggest an increase in automaticity. In terms of second language learning, the CV_{RT} could also indicate the similarities and differences between second language performance (qualitative processing) and native language performance (Segalowitz et al., 1998). In Experiment 1, the sleep group also experienced smaller CV_{RTs} than the wake group for the Malay words. When comparing CV_{RTs} between Malay and English words, the wake group had greater CV_{RTs} for Malay compared to English words. In contrast, there was no significant difference in CV_{RTs} between Malay and English words for the sleep group, suggesting that after a night's sleep (but not an equivalent duration of wakefulness), participants' variations in RT for the new words were similar to their variations for existing English words that were consolidated and integrated in memory. These results further support the notion that sleep plays a global role for achieving memory consolidation and integration of new words with greater consistency and stability in responses to the newly learnt words. The smaller CV_{RTs} exhibited by the sleep group compared to the wake group also suggest that sleep enables greater automaticity in processing newly learnt words, which lends support to the notion mentioned previously that the size congruity and semantic distance effects were representative of the automaticity with which word meanings are assessed.

Even though the participants in the study were young adults, the second language learning paradigm enabled parallels to be drawn with development research. When considering the size congruity and semantic distance effects across development, fMRI studies have shown that when completing tasks investigating the distance effect in numerals, different brain areas were activated in children compared to adults, where semantic distance effects were associated mainly with frontal regions in children and parietal regions in adults (Ansari, Garcia, Lucas, Hamon, & Dhital, 2005; Kaufmann et al., 2006). It has also been found that the above differences can be attributed to greater automaticity in processing the relationship between the individual numeral symbols and their semantic magnitudes throughout development (Ansari et al., 2005; Ansari, 2008). Gebuis, Cohen Kadosh, de Haan, and Henik (2009) investigated size congruity effects for Arabic numerals in 5 year-old children and adults. Unlike the adults, the children did not show a size congruity effect for the Arabic numerals. In contrast, Zhou et al., (2007) found a size congruity effect in Chinese children of a similar ages. The experimenters mentioned that the Chinese children studied were exposed to the semantic relationship between numerals in their everyday knowledge at an earlier age than Western children. These findings imply that the size congruity effect may occur later in development when the newly learnt information is integrated with existing knowledge that is relevant to everyday life. Since the sleep group was the only group that displayed a size congruity effect for the new second language words in Experiment 1, it is plausible that sleep plays an active and direct role in integrating the newly consolidated words with existing (mother tongue) word knowledge, as participants in the sleep group were unable to ignore the physical font sizes of the new Malay words where they did not match their referent semantic meaning in relation to existing knowledge. The present findings fit well with existing developmental research. The above studies show that children develop stronger size congruity and semantic distance effects across development, resulting from differences in brain activation and levels of knowledge integration. Therefore, sleep-associated behavioural changes in the current study may underlie the above changes. The current findings suggest that the behavioural changes in size congruity and semantic distance effects after sleep are markers of change in processing the new words after sleep. These markers of change in processing may be linked with neural changes during sleep, leading to the integration of the new words into existing knowledge.

Moreover, even though researchers who have studied the size congruity and semantic distance effects have linked both effects to the automaticity with which meanings are assessed (see Tzelgov, 1999 for a review), it is likely that the size congruity effect is a stronger demonstration of automatic semantic access than the semantic distance effect. The above suggestion stems from stimulus-response (S-R) compatibility (SRC) models (Sanders & McCormick, 1976). The term SRC is used to denote the notion that task difficulty may be due to individual task stimuli, response needed or the relationship between the combined S-R mappings

(Kornblum, Hasbroucq & Osman, 1990). According to the dimensional overlap SRC model by Kornblum et al (1990) the SRC effect occurs when there are shared or similar properties between task stimuli and response regardless of whether the shared properties are related to task instructions/demands (Eimer, 1995). When there is consistency between the S-R mappings, participants' RTs are predicted to be swifter as the correct response would match what has already been highlighted in the task instructions and hence according to the experimenters, be automatically activated. In contrast, when the S-R mappings are inconsistent, participants' RT would be slower as participants' need to first inhibit the response that is automatically activated before selecting the correct response (Eimer, 1995; Kornblum et al., 1990).

When thinking of the semantic distance effect in terms of consistency between S-R mappings, one may view the semantic distance effect to be related to fully consistent S-R mappings as there is no conflict between the stimulus word-pair and the task instruction to select the semantically larger item. In contrast for the size congruity effect, consistency between S-R mappings is affected by the congruity between the word-pairs. When the word-pair is congruent, there is consistency between S-R mappings. However, for the incongruent word-pair, participants need to inhibition the mismatch between font and referent size in order to correct select the semantically larger item. For the newly learnt words, the presence of inhibition (slower RTs) for incongruent trials in the size congruity effect would indicate that meanings are automatically accessed to the threshold such that interference occurs when the mismatch between physical font- referent mappings affects participants' response to select the item. Therefore, for the new word stimuli, it is predicted that the emergence of the size congruity effect would require stronger S-R mappings as compared to the semantic distance effect.

Applying findings from the current study to a CLS consolidation model of word learning (Davis & Gaskell, 2009; see also general systems consolidation models in Frankland & Bontempi, 2005; Walker, 2009) when acquiring a new language, the form-meaning links for each new word are rapidly encoded by the hippocampal system during initial exposure as compared to the slow-learning neocortical system. In this model, sleep is required for the gradual reliance of individual form-meaning mappings from the 'fast-learning' hippocampal to the 'slow-learning' neocortical system to enable more automatic processing of the representations. Viewing this in relation to the subtypes of automaticity described above, the semantic distance effect, which is thought to reflect automaticity emerging at a lower threshold (weaker S-R mappings), was present in both the sleep and wake groups. However, the size congruity effect which thought to reflect a stronger measure of automatic semantic access which requires stronger S-R mappings, was present only in the sleep group for the new Malay words (in the cases where interactions between group and congruity were present). Therefore, it is likely that the semantic distance effect draws more on the strength of specific individual mappings whereas the size congruity effect draws mostly on general and global information integrated across existing knowledge. This also suggests that the consolidation or strengthening of individual form-meaning mappings occurs during wakefulness but is further strengthened during sleep. In contrast, the integration of the individual mapping with existing knowledge may be primarily sleep-dependent.

By examining sleep polysomnography data, Experiment 1 also aimed to explore if there was a direct relationship between the effects of sleep on integration in word learning. It was predicted that both sleep spindle and SWS activity would play an additive role in consolidation and integration of the new Malay words. This prediction was supported, as participants who experienced greater amounts of SWS also displayed stronger semantic distance effects in Malay word-pair comparisons (congruent trials). Similar to the behavioural findings, there was no relationship between SWS and the strength of the semantic distance effect for the comparable trials when looking at existing English word-pairs, suggesting that the above finding was not confounded by individual sleep quality. In addition, there was a positive correlation between NREM sleep spindle count and behavioural findings of the strength of the size congruity effect for new Malay but not existing English word-pair comparisons with larger physical size differences and semantic size distances. SWS may play a main role in strengthening individual episodic form-meaning mappings in the neocortex, whereby reactivation of hippocampal memories during SWS enables the individual mapping to be more reliant on the neocortical system for storage and retrieval. In comparison, sleep spindles and their interplay with slow oscillations and sharp-wave ripples may be more crucial for the integration of these individual mappings with existing knowledge. Although further research is needed for a more comprehensive view, these findings give additional insights on how specific components of sleep might operate within the CLS model.

Aside from the CLS model, the current findings also give insight into other general systems consolidation models of sleep-associated memory consolidation, such as the model of Walker and Stickgold (2010). Their model can be summarised by three key components termed by the authors as ‘unitization’, ‘assimilation’ and 'abstraction'. The process of 'unitization' involves the strengthening of the new cortical representations. The process of 'assimilation' can be paralleled with the integration of new neocortical mappings with existing knowledge in the CLS. In addition to integrating new memories with existing knowledge, Walker and Stickgold's model also attributes a crucial role of sleep to the 'abstraction' of knowledge, where sleep-dependent integration further enables the generalisation of knowledge such that novel relationships between the new unitised and assimilated representations are formed. The current findings concur with the model. Participants in the sleep group were able to form a generalised hierarchy (small < medium < large animal) regarding the size of the animals that the new words referred to, as the processing of new form-meaning mappings in Malay were affected by the level of congruency in relation to existing knowledge, for participants who slept after learning. Moreover, the differences in actual (real-life) animal sizes affected the processing of new word-pairs (semantic distance effect), implying that novel associations that were not explicitly taught during training were formed between the newly learnt words. With a second language learning paradigm, the current findings also suggest that sleep enables strengthening (unitisation) of the newly integrated form-meaning representations with semantic knowledge in existing English words in the neocortical system (assimilation), whereby shared properties (about animal sizes) among representations were extracted, allowing participants to form generalised associations between the new representations (abstraction).

In summary, Experiment 1 investigated the effects of sleep and specific components of sleep on the integration of new knowledge, by comparing differences in size congruity and semantic distance effects for newly learnt and existing English words in participants who either slept after learning or remained awake. Behavioural findings provided initial evidence for an association between sleep and the integration of new word meanings: when making semantic comparisons of newly learnt words, size congruity and distance effects were stronger after a period of sleep compared to wakefulness. By incorporating sleep polysomnography, Experiment 1 aimed to investigate whether sleep also played a more causal and direct role on declarative knowledge integration. Results indicated that individual sleep components were associated with integrating meanings in new words; SWS was related to the strength of the distance effect and spindle count was related to the strength of the congruity effect. The overall findings were in line with a systems consolidation account of an active role of sleep on declarative memory consolidation and integration. In regards to the role of sleep in word learning, the findings suggest that individual form-meaning mappings learnt during wakefulness are further strengthened during sleep and that the integration of the individual mappings with existing knowledge may be primarily sleep-dependent.

Chapter 3: Effects of Day-time Napping and Variations in Training on Knowledge Integration in New Word learning

3.1 Introduction

Previous work has highlighted that in addition to overnight sleep, sleep-associated benefits in declarative memory can also occur during a nap. For example in the study by Lau, Alger, and Fishbein (2011) cited in Chapter 1[6], participants were able to abstract untrained rules in new Mandarin words after a 90 minute nap but not after an equivalent period of wakefulness. Mednick, Cai, Kanady and Drummond (2008) also investigated the effects of daytime napping compared against caffeine and placebo on a verbal word list recall and recognition task, a procedural finger tapping task and a perceptual textile discrimination task. Napping, compared against both caffeine and placebo, led to significantly better recall and recognition of words for the verbal task after a 7 hour training-test interval. In contrast, there were no nap-associated benefits for the finger tapping task and no significant difference in the benefits of napping and caffeine for the textile discrimination task. The above findings suggest that naps seem to have a specific benefit for declarative memory as compared to other forms of memory.

Still, it would be useful to see if differences in behavioural performance for participants who napped would be also related to differences in their sleep

[6] Lau, Alger and Fishbein (2011), participants learnt 7 sets of Mandarin characters and their relevant English meanings. Each set of characters contained a common property known as radicals, which are often used to inform Mandarin readers about the meaning of the characters. After either a 90 minute nap or an equivalent period of wakefulness, participants were given tasks containing untrained Mandarin characters with the same radicals as the sets of trained characters. For each untrained Mandarin character, there were four possible options for its correct meaning. In order to find/select the correct answer for this task, participants had to abstract commonalities between the untrained character and the corresponding set of trained characters with the same radical. For the next task, participants were also shown individual radicals and had to indicate what they thought the radicals meant. Participants who had taken a nap between training and test achieved significantly higher accuracy compared to those who remained awake for both tasks. Hence, findings indicate that sleep is central in integrating information across meanings of new words, enabling abstraction of previously untrained semantic rules between words.

architecture. This would attribute a more direct association of the effects of daytime napping on declarative memory consolidation. Also mentioned in Chapter 1, Lau, Tucker, and Fishbein (2010) found in their study of declarative memory that participants who napped between training and testing remembered significantly more face-face associations when tested 2 hours after training compared to participants who remained awake. Most importantly, improvements in performance were significantly positively correlated with the amount of SWS participants experienced. The above findings suggest that SWS during naps play an active role towards the consolidation of new declarative information. Studies have also shown that post-training naps can lead to long-term benefits in declarative memory consolidation. Takashima et al., (2006) investigated the effects of sleep on declarative visual recognition memory for an extended time course where participants were tested after learning at days 1 (same day as learning), 2, 30 and 90. Participants were given a 90 minute nap opportunity between training and testing on day 1, where the nap was monitored with polysomnography to directly investigate effects of sleep architecture on subsequent behavioural performance. There was a positive relationship between SWS experienced during the day 1 nap and recognition performance not only on day 1 but also on day 2 and day 30. In contrast, there was no equivalent relationship with any other sleep stage, suggesting that SWS is directly related to the memory consolidation of new declarative information.

Moreover, a separate study reported benefits for declarative memory even with shorter (post-learning) nap duration. Lahl, Wispel, Willigens and Pietrowsky (2008) reported that after learning a list of 30 words, a nap that lasted approximately 6 minutes was sufficient to promote better recall of the words as compared to wakefulness. The predominant sleep stage in the 6-minute nap was stage 1 sleep. None of the participants had SWS. This result was interesting especially in relation to system consolidation models which place a great emphasis on slow oscillations (mainly in SWS) and sleep spindle activity (mainly in stage 2 sleep). Still, compared to a subsequent study by the same experimenters where participants napped for approximately 35 minutes, participants who had the longer 35-minute nap recalled significantly more words than those in the shorter 6-minute nap study. Apart from having a longer amount of total sleep time (TST), all but 2 of the participants in the longer nap group entered SWS. In addition, even though the relationship did not reach significance, the amount of SWS was positively related to recall performance. This suggests that the enhanced performance in the longer nap group could be due to time spent in SWS in addition to longer TST. The dissociation between the effects of TST and SWS on declarative memory was proposed by an earlier nap study by Schabus, Hödlmoser, Pecherstorfer and Klösch (2005). Participants were taught a list of word-pairs and memory for the word-pairs was tested before and after a 60-minute nap opportunity. A positive correlation was found between improvements in memory performance (before and after napping) and TST during the nap. However, post-hoc tests revealed that participants who experienced SWS during their nap had significantly better memory performance in the post-nap compared to pre-nap session. In contrast, there was no significant improvement in memory performance between pre and post-nap sessions for participants who did not experience SWS. Therefore, even though greater TST is related to better declarative learning, SWS seems to have a specific benefit towards declarative memory consolidation.

3.2 Experiment 2

Although a nap design may be useful in investigating effects on sleep on declarative memory consolidation, it is important to consider the potential weaknesses of a nap design. One potential weakness with a nap design is the large variability in sleep quality experienced by participants. For example, in the longer nap study by Lahl et al. (2008), there was also large variation in the key sleep parameter investigated, whereby the standard deviation for time in SWS (9 minutes) was very similar to the mean time spent in SWS (10 minutes). It is difficult to control for these variations and it is possible that the larger variability in nap studies make them a less sensitive measure in predicting sleep-associated effects. In clinical studies comparing nap and overnight polysomnography, nap studies were viewed as poorer predictors of sleep abnormalities than overnight studies. Clinicians have also highlighted that findings from nap studies may not be transferrable to results from overnight studies (Saeed, Keens, Stabile, Bolokowicz, & Davidson Ward, 2000). However, there are also strengths that a nap polysomnography study design has over an overnight study. One of the limitations of Experiment 1 which adopted an overnight design was that participants in the wake and sleep groups were trained and tested at different timings. Even with the English baseline trials, it was not possible to fully eliminate the possibility that items learnt in the morning and tested in the evening (wake group) would be retrieved differently from items learnt in the evening and tested in the morning. A nap study would be able to fully control for any possible time of day effects as participants in the wake and nap group will be trained and tested at the same timings, making sleep the only key variable that differs between the experimental groups.

In a recent study, Alger, Lau and Fishbein (2012) investigated the effects of sleep, TST and SWS on declarative memory consolidation by comparing participants who experienced one of three conditions after a word pair-associate learning task. Participants experienced one of three conditions during the 75 minute training-test interval: a 10-minute nap without SWS, a 60-minute nap with SWS or remaining awake. At the first test point, participants in the 60-minute nap condition remembered significantly more words than those who remained awake. Participants who had a 10-minute nap had marginally better performance than the wake participants. At the second test point after an interference task, participants in the 60-minute nap condition recalled significantly more word-pairs than all other groups. A third test was conducted after one week and results mirrored that of the second test point where the 60-minute nap group had better performance compared to the 10-minute nap group and wake group. Alger, Lau, and Fishbein (2012) interpreted results in the first test point as evidence for a benefit of sleep for declarative learning and results in the second test point as evidence that the quantity of sleep or TST was important towards benefiting declarative memory consolidation. In addition, the experimenters predicted that sleep played an active role in declarative consolidation, where the additional SWS[7] experienced by the 60-minute nap group lead to long term sleep-associated gains in memory consolidation exhibited in the third test point. Even though the experimenters looked at variations in TST and individual sleep stages between different nap groups, their findings suggest that these variations may be interesting to investigate as individual differences within a nap period. Therefore the individual differences in quantity and quality of sleep during a nap period may be useful may considering individual differences in sleep-associated gains for declarative memory consolidation.

Drawing from previous research where a nap period was related to enhanced memory consolidation, Experiment 2 aimed to investigate the effects of daytime napping compared to wakefulness on knowledge integration in new word learning using size congruity and semantic distance effects. Unlike Experiment 1 where participants were allocated to either the wake or sleep group, participants in Experiment 2 took part in both the wake and nap condition in order to further reduce individual differences and more directly attribute disparities in the wake and nap conditions directly to sleep. It was hypothesized that participants would exhibit

[7] There was no significant difference in performance when comparing participants who experienced REM sleep with those who did not.

stronger size congruity and semantic effects for the newly learnt words in the condition where they experienced a nap after training as compared to remaining awake. It was also predicted that there would be no equivalent differences in participants' performances in the nap and wake conditions for baseline English animal names. This would attribute the behavioural benefits for the new words more directly to sleep. Based on studies mentioned in section 3.1 (Alger et al., 2012; Lahl et al., 2008; Schabus et al., 2005) and results from Experiment 1, Experiment 2 also aimed to explore the effects of TST, SWS and sleep spindle activity (Experiment 1) on changes in automaticity for new word learning. Unlike the work on SWS, there have been limited nap studies looking at the relationship between sleep spindles and declarative learning. Still, drawing from findings in Experiment 1, it was predicted that the strength of the size congruity and semantic distance effects for the newly learnt words would be positively related to sleep spindle activity and SWS (Alger et al., 2012; Lahl et al., 2008; Schabus et al., 2005) respectively. It was also predicted that participants who experience greater TST would show stronger gains in automaticity than those with less TST (Alger et al., 2012).

Linking nap-associated consolidation to a systems consolidation model, Alger et al. (2012) suggested that it was possible that participants in the 60-minute nap condition experienced greater (and longer-term) memory gains due to sleep-dependent reactivation which may have occurred in the additional SWS the participants experienced. When examining the complementary learning systems model of word learning (CLS; Davis & Gaskell, 2009), it may be possible that the transfer of memory strength from the hippocampal to neocortical networks can occur during a nap period but may be qualitatively different to overnight consolidation. By comparing findings from Experiment 1 (overnight study) and the current Experiment 2 nap study, it may be possible to also gain further insights to the time course and the effects of sleep architecture on systems consolidation models of declarative learning.

In addition, behavioural findings from Experiment 1 revealed that participants who slept after learning had similar coefficients of variation in RT (CV_{RT}) for Malay and English words, whereas participants in the wake group had significantly larger CV_{RTs} for Malay compared to English words. Since CV_{RT} was proposed to be a marker of automatic processing (Segalowitz & Segalowitz, 1993; Segalowitz, Segalowitz, & Wood, 1998), results implied that overnight sleep (but not an equivalent period of wakefulness) leads to enhanced automaticity in processing newly learnt words that matched well-integrated words. Experiment 2 aimed to further investigate sleep-associated changes in CV_{RT} for word learning and examine if a period of nap was sufficient to promote enhanced automaticity of processing in newly learnt words.

3.2.1 Methods

Apparatus and Materials

The test apparatus consisted of a computer with a 15-inch screen. DMDX software (Forster & Forster, 2003) was used to collect all RT and error data. Participants used the USB-joypad for all experimental responses. As participants in Experiment 2 took part in both the nap and wake conditions, a larger stimuli set was needed compared to Experiment 1. It should also be noted that Malay words in Experiment 1 were direct translations of the English animal names they were trained with, but due to the larger stimuli set and changes in stimuli selection criterion in Experiment 2, it was not possible to train participants with direct Malay-English translations.

A total of 18 Malay words and 36 English animal names were selected for the whole experiment (see Appendix 4) whereby the experimental stimuli for each condition consisted of 9 Malay words and 18 existing English animal names. The Malay words were selected from the ‘Malay Lexicon Project’ database (Yap et al., 2010) where all 5-letter Malay bi-syllabic words with no orthographic neighbours, were ranked in terms of frequency using a English-Malay bilingual population. 36 Malay words that had the highest frequency based on the above criteria were each rated based on their ‘pronounceability’ (i.e. how easy it was for participants to pronounce the word) on a scale of 1 (very easy to pronounce) to 7 (very difficult to pronounce) by 20 native British English speakers. After ‘pronounceability’ ratings were collected (M = 2.13, SD = 1.16), 18 Malay words that were rated easy to pronounce by majority of the raters (M = 1.82, SD = 0.27) selected as experimental stimuli.

Animal names were chosen from a list of 60 animals taken from mono- and bi-syllabic animal names from Paivio's (1975) size norm ratings and a popular children’s learning website[8]. 23 raters ranked the 60 animals on physical size using a scale of 1 (smallest) to 7 (largest). Lexical decision RTs (M = 628, SD = 70) from the English Lexicon Project database (Balota et al., 2007) and frequency ratings (M = 8.21, SD = 1.25) were calculated from the MRC Psycholinguistics database

[8] (http://www.enchantedlearning.com/subjects/animals/Animalbabies.shtml)

(Coltheart, 1981). 36 animal names with word lengths between 3 to 6 letters were selected as experimental stimuli: all selected stimuli also had lexical decision RTs and familiarity ratings within 2 standard deviations from the means of the 60 animal names. The selected animal names were ranked according to mean size norms and allocated to 3 size groups: Small (*M* = 1.07, *SD* = 0.46), Medium (*M* = 2.43, *SD* = 0.72) and Large (*M* = 5.26, *SD* = 0.83), with 12 animals in each group. Items in each size group were matched whereby an ANOVA revealed no significant differences in lexical decision RTs, frequency ratings and word length between groups (all p's >.27). Items in each size group were then randomly distributed into 4 stimuli sets and each animal name was randomly paired with one Malay word. 36 photographs measuring 45x30mm were selected from Google-images to depict each animal in all sets. Photographs were matched whereby the animal took up approximately ¾ of the total area and all backgrounds were neutral (natural habitat).

Similar to Experiment 1, a measure of general sleep quality was taken using the Pittsburgh Sleep Quality Index (PSQI). The Stanford Sleepiness Scale (SSS) was also administered to indicate how sleepy or alert participants felt.

Stimuli & Design

For the training session, the exposure and feedback trials were manipulated in the same way as Experiment 1. For the test session, all possible pairs of stimuli were generated for the Malay and English animal names. For semantic size distance, items differed by either small (small-medium, medium-large animal) or large (small-large animal) distances. Unlike Experiment 1, for physical size difference, items only differed in font height 4mm (one item was 7mm and the other 11mm). The item-pairs also had a centre-to-centre distance of 50mm. As in Experiment 1, there were 18 possible item-pairs differing by small semantic size distances and 9 item-pairs differing by large semantic size distances for each stimulus set presented in the same language (English/Malay). Congruity was manipulated whereby half the trials were congruent and half were incongruent. Each possible pair was then presented twice, counterbalancing items on the left and right to prevent biasness in response. To ensure that participants viewed the same number of small semantic size distance and large semantic size distance pairs, each counterbalanced large semantic size distance pair was then presented twice. Moreover, to investigate potential confounds such as further 'learning' during the test session or fatigue effects due to large number of trials; the full set of stimuli was repeated twice (block of presentation) in Experiment 2. Therefore, for each comparison task, 288 trials were presented in a random order for each participant: 36 (small and large semantic size distance-pairs) x 2 (left/right counterbalancing) x 2 (congruity) x 2 (block of presentation); for a total of 576 trials[9] for both English and Malay comparison tasks combined.

In summary, a mixed-subjects design was used. The dependent variable was RTs for correct responses. There were five within-subjects independent variables, namely: experimental condition (wake/nap), language (English/Malay), congruity (congruent/incongruent), semantic size distance (small/large) and block of presentation (first/second presentation). The between subject variable was session sequence (wake session first or nap session first). A separate analysis was also conducted with proportion of errors as a dependent variable and is presented in Appendix 1.2.

Participants

24 monolingual, non-smoking, native English speakers (14 males, 10 females; *M* = 21, range: 18-30) with no known language disabilities and no prior history of drug or alcohol abuse, neurological, psychiatric or sleep disorders (screened by the PSQI) were recruited via an online database and were given a cash reward or course-credit for their participation. Participants were randomly allocated to one of the four counterbalanced groupings (n = 6 per group). All participants had a sleep-wake pattern where they would typically rise by at the latest by 9am each morning and were used to having at least 6 hours of sleep per night.

Participants were also emailed a with 'participant information form' before the experiment. They were instructed prior to the training session to have a night of normal sleep (based on their usual sleeping patterns). In addition, they were required not to consume caffeine throughout the experiment and to refrain from taking recreational drugs or alcohol 24 hours prior to the training sessions and also throughout the training-test interval. The study was approved by the Department of Psychology (University of York) ethics committee.

[9] The number of test trials was identical to Experiment 1 because even though there was only one level of physical font size manipulation, the new variable of block of presentation was introduced in Experiment 2.

Procedure

Testing procedure was adapted from previous research on sleep and memory consolidation using nap studies with a 90 minute nap opportunity (Backhaus & Junghanns, 2006; Nishida & Walker, 2007; Takashima et al., 2006, see Figure 6). Each participant underwent 2 training-test sessions carried out in the University of York Psychology departmental sleep laboratory. Participants first arrived in the laboratory at approximately 12:00pm for training and remained till approximately 4:00pm for testing. At the start of the first session, participants filled in a participant consent form, the Pittsburgh Sleep Quality Index (PSQI) and the Stanford Sleepiness Scale (SSS). They also filled in another SSS at the start of each subsequent test or training session. Prior to the training session, all participants underwent an electrode setup for polysomnographic recording to prevent them from predicting which condition (wake/nap) they were allocated to. Throughout training and testing, participants were seated at approximately 60cm viewing distance from the screen. Between initial training and subsequent test, participants experienced either a 90-minute nap opportunity with sleep polysomnography recording (nap condition) or an equivalent period of rested wakefulness (wake condition). Therefore, the main difference between the conditions was whether participants took a nap or remained awake after the training session. Participants took part in both nap and wake conditions, separated by a 1-week interval (order of conditions was counterbalanced).

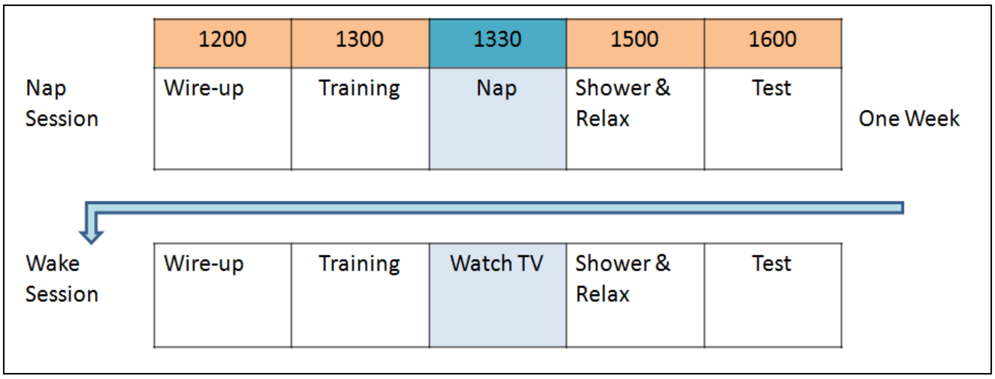


Figure 6. Experiment 2: Experimental procedure for participants

Training. The procedure for the training session replicated that of Experiment 1.

Test. Participants carried out four comparison tasks in a test session: two with known English words to control for circadian effects and two with the Malay words. The English tasks consisted of animal names from the opposing stimuli set and the Malay tasks consisted of words that participants learnt in the training session. Participants always completed one English task first, followed by one Malay task. Next, the procedure was repeated (for a second block). The trials within the two English tasks and the two Malay tasks were identical, but were presented in different randomized orders within each task. All remaining procedure in the test session was similar to Experiment 1.

In addition, at the end of the whole experiment, the Wide Range Achievement Test (WRAT-3) (Wilkinson, 1993) was administered. The WRAT-3 test comprised of spelling, reading and arithmetic components. As there were no explicit numeral components in Experiment 2, only the spelling and reading components were analysed. It should be noted that the vocabulary and matrix reasoning subtests from the Wechsler Abbreviated Scales of Intelligence (WASI) (Wechsler, 1999) that were used in Experiment 1 are thought to be independent of factors such as education and a test of general intelligence, whereas the WRAT-3 measures levels of achievements in factors that may be determined by education. The rationale for using the WRAT-3 was that it separated spelling and reading ability which may be useful measures to control for.

Polysomnographic (PSG) recording. In Experiment 2, electrodes were set up following the International 10-20 system as recommended by the American Academy of Sleep Science Manual (AASM, 2007). A 13 channel montage was used: with 6 scalp electrodes referenced to contra-lateral mastoids (F3-M2, F4-M1, C3-M2, C4-M1, O1-M2, O2-M1), left and right electro-oculographic (EOG) channels and 3 chin electromyographic (EMG) channels.

EEG Analysis. All technical and digital specifications, including PSG impedance levels and sampling rates and filter settings were set according to the recommended specifications in the AASM. Data collected was also scored manually in 30 second epochs according to the AASM sleep staging criteria. Channels that were noisy or had poor quality signal for majority of the recording were removed from scoring analysis. In addition, 20% of the data was cross-scored by a second independent scorer. A computer generated inter-scorer reliability report revealed an average of approximately 85% overall score agreement between the 2 independent scorers. Sleep spindles were analysed using the same methods as Experiment 1, with the change that the frontal electrodes were used instead of the central ones, as the AASM (Iber, Ancoli-Israel, Quan, 2007) highlighted that spindle activity is greatest in the frontal regions.

3.2.2 Results

Data from 3 participants were excluded as they either remained awake for more than half their 90 minute nap opportunity (n = 1) or did not achieve at least 50% of NREM (stage 2 and SWS) sleep (n = 2). There were no significant differences in Stanford Sleepiness Scale ratings between training and testing in both wake and nap conditions for the remaining 21 participants (all *p*'s > .16). Box plot analysis of WRAT-3 reading and spelling scores also revealed that all participants were within normal distribution for their age and educational level.

Training Session

Table 6. Experiment 2: Mean response time (in milliseconds) and error rates (percentage) for the training session. Standard errors are presented in parentheses.

	Block 1		**Block 2**	
	Wake	Nap	Wake	Nap
RT	1147 (365)	1146 (337)	1086 (334)	1071 (315)
Error Rate	6.6 (4.4)	6.5 (4.3)	5.4 (3.3)	6.1 (3.5)

Performance in the training session was separated into two blocks (separated by the untimed break) and presented in Table 6. As the mean error rate across participants for the whole training phase was 6.1%, close to twice of that in Experiment 1, participants' training data were also further analysed as a covariate in the test phase, to evaluate if performance during the training phase would affect subsequent test performance (see analysis in the Test session).

Test Session

Table 7. Experiment 2: Mean response time (in milliseconds) for all correct responses after data trimming. Standard errors are presented in parentheses.

	English		**Malay**	
	Wake	**Nap**	**Wake**	**Nap**
Block 1				
Congruity				
Congruent	804 (39)	820 (32)	1306 (87)	1268 (58)
Incongruent	819 (38)	854 (35)	1367 (73)	1291 (60)
Semantic Distance				
Small	862 (40)	905 (37)	1403 (82)	1413 (65)
Large	762 (37)	770 (31)	1209 (76)	1146 (53)
Block 2				
Congruity				
Congruent	818 (57)	758 (35)	1028 (53)	999 (50)
Incongruent	829 (45)	769 (34)	1044 (54)	1003 (49)
Semantic Distance				
Small	831 (42)	893 (53)	1096 (55)	1127 (58)
Large	697 (27)	754 (49)	905 (45)	945 (48)

RTs underwent two trimming procedures. Identical to Experiment 1, mean RTs for all correct responses were separated into congruent and incongruent trials within the Malay and English comparison tasks. Initial analysis also revealed a main effect of presentation block where participants were significantly faster at responding in the first block of presentation as compared to the second for both Malay [$F(1, 20) = 66.13$, $p < .001$] and English [$F(1, 20) = 4.98$, $p = .037$] comparisons. Therefore, in addition to the trimming procedures adopted from Experiment 1, mean RTs for the above trials were trimmed separately based on presentation block (first vs. second). Hence, trials more than 2 standard deviations away from each participant's mean RT for the corresponding task were discarded and the remaining RTs were presented in Table 7. This resulted in an average data loss of 6.2% and 5.6% per participant for the Malay and English comparisons respectively. Next, participant's mean RT for the second block of training phase (RTs above 2000ms were excluded from this analysis) was combined for both wake and nap condition and used as a covariate for the ANOVA analysis. The second training block was selected as it was thought to be more representative of final training performance (participants had smaller deviations in error rates).

Prior to the ANCOVA analysis, each participant's mean training RT was centred using the Delaney-Maxwell method by subtracting the mean training RT across participants (Delaney & Maxwell, 1981). Therefore, a negative mean-centred RT indicated that the participant had quicker than average training performance and a positive RT indicated slower than average performance. The centring of training RTs was done to ensure that there were no changes in main effects after carrying out the ANCOVA (Algina, 1982). A seven-way mixed ANCOVA was performed with session sequence (wake condition first or nap condition first) as the between-participant factor, language (English/Malay), condition (wake/nap), presentation block, congruity and semantic distance as within-participant factors and training performance as a covariate. In general, participants who performed with swifter RTs in the training task also had quicker overall for both the English [$F(1, 18) = 10.22$, $p = .005$] and Malay test trials [$F(1, 18) = 30.77$, $p < .001$]. However, it is important to note that the training covariate was also related to learning of new items rather than baseline RT performance as there was an interaction between language and training [$F(1, 18) = 6.01$, $p = .025$], where the relationship between training and test RTs was greater for the newly learnt Malay words as compared to English words. There was a large main effect of language in the ANCOVA [$F(1, 18) = 84.56$, $p < .001$], where participants were significantly faster at responding to English as compared to Malay word-pairs. Hence in the cases where main effects and interactions were significant for the combined (English and Malay) ANCOVA, further analysis was conducted to investigate if the main effect and interactions were present in both languages or only one language. As mentioned in Experiment 1, this was to ensure that results were not affected by the substantial difference in absolute RTs between languages. Interactions with training covariate were only reported if they were different from the interactions without the training covariate. In addition, interactions with presentation block were not reported except for cases of theoretical interest.

The semantic distance effect was significant in the combined ANCOVA [$F(1, 18) = 248.07$, $p < .001$] and also in both the English [$F(1, 18) = 125.51$, $p < .001$] and the Malay [$F(1, 18) = 203.82$, $p < .001$] trials (see Figure 7a). The size congruity effect was significant in the combined ANCOVA [$F(1, 18) = 5.11$, p = .036], but when looking at the languages separately, the size congruity effect was present in the English [$F(1, 18) = 5.08$, p = .037] but not the Malay trials [$F(1, 18)$ = 1.53, p = .23] (see Figure 7b).

(a)

(b)

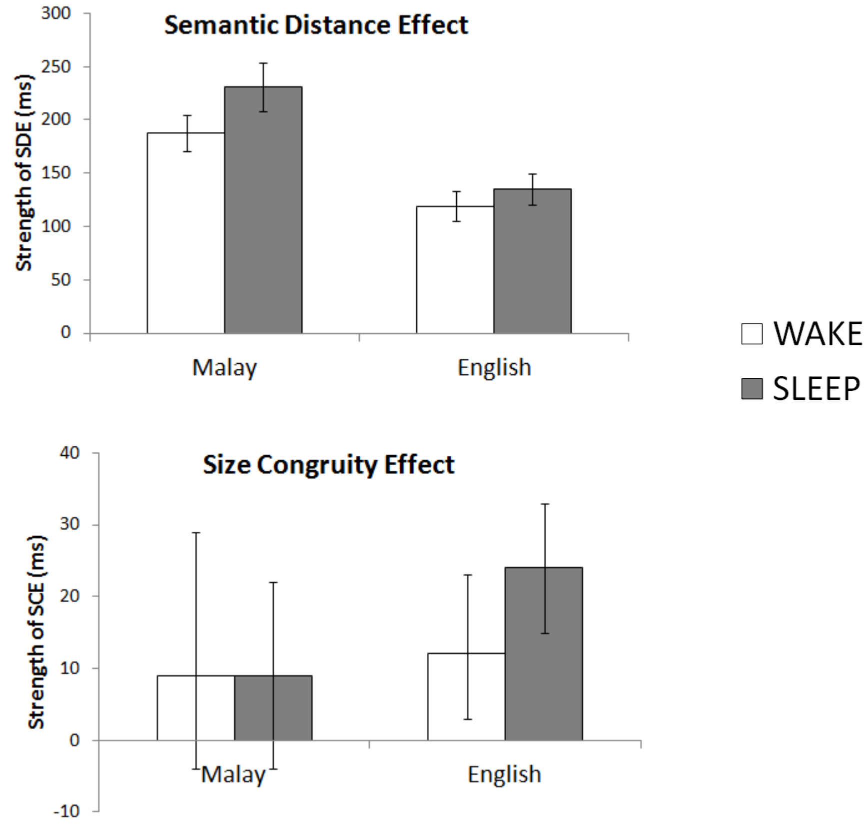


Figure 7. Experiment 2: Strength of Automaticity Effects: (a) Semantic Distance Effect (difference in RTs between trials with small distances and large distances), (b) Size Congruity Effect (difference in RTs between incongruent and congruent trials). Error bars

There was also a significant four-way interaction between language, condition, semantic distance and training [$F(1, 18) = 10.00$, p = .005]. When looking the language trials separately, there was a marginally significant interaction between condition, semantic distance and training performance [$F(1, 18) = 4.49$, p .048] for the Malay trials but not for the English trials [$F(1, 18)$ = .68, p = .42].

As participants took part in both the nap and wake condition, it was possible to further analyse four-way interaction by looking at the difference in strength of semantic distance effect between the nap and wake condition (i.e. deducting the strength of semantic distance effect during nap from strength of semantic distance effect during wakefulness) for both the Malay and English trials. Correlations revealed that there was a significantly positive relationship between nap-associated gains in strength of semantic distance and training performance for the Malay word-pairs, *r* = .44, *p* = .046 (see Figure 8a). In contrast, there was no equivalent relationship between nap-associated gains in strength of semantic distance and training performance for the English word-pairs, *r* = -.19, *p* = .41 (see Figure 8b). Moreover, a Steiger's z-transformation affirmed that the nap-associated benefit to the Malay words was significantly larger than that for the existing English words, *z* = 2.49, *p* = .013.

(a) (b)

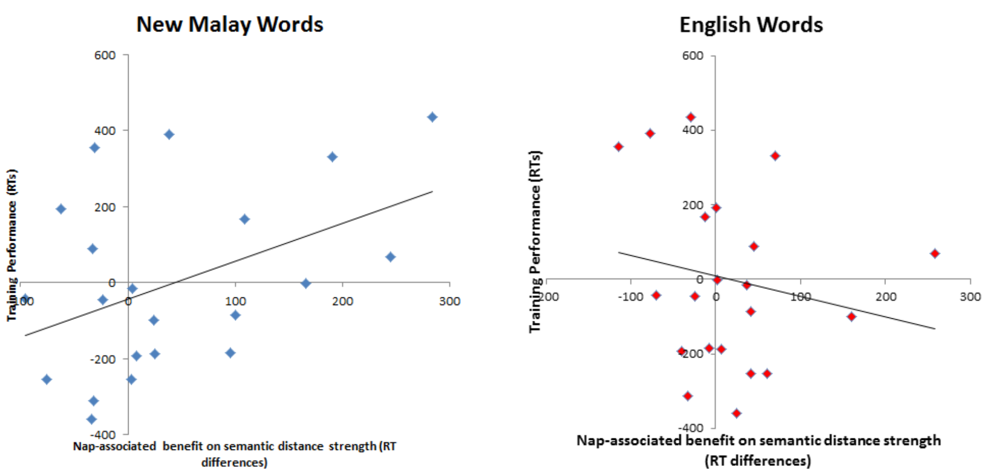


Figure 8. Experiment 2: Correlation between training performance and nap-associated gains in strength of semantic distance effects for (a) Malay words and (b) English words

There was also a marginally significant language by block, condition, congruity and the training covariate interaction [*F*(1, 18) = 3.29, *p* = .086]. The interaction between block, condition, congruity and training was significant across the Malay [*F*(1, 18) = 5.28, *p* = .034] but not the English [*F*(1, 18) = .29, *p* =.60] trials. Further analysis of the Malay trials indicate that the interaction between condition, congruity and the training covariate was significant in presentation block 1[10] [*F*(1, 18) = 5.48, *p* = .031] but not block 2 [*F*(1, 18) = .18, *p* = .68]. Next, the difference in strength of size congruity effect between the nap and wake condition for both the Malay and English block 1 trials, was analysed using the identical method as the semantic distance effect. There was a significant positive correlation

[10] A language by condition, congruity and the training performance analysis of all block 1 trials also indicated that the interaction was greater in the Malay compared to English trials [*F*(1, 18) = 5.20, *p* = .035]

between nap-associated gains in strength of size congruity (block 1) and training performance for the Malay word-pairs, $r = .49$, $p = .024$ (see Figure 9a), but no equivalent relationship for the English word-pairs, $r = .03$, $p = .91$ (see Figure 9b). However, visual examination of Figure 9a suggests that the data may deviate from normality, whereby when removing the right-most data point in Figure 9a, the correlation between nap-associated gains in strength of size congruity (block 1) and training performance for the Malay word-pairs was only marginally significant, $r = .40$, $p = .084$. Therefore, in order to preserve the full dataset, non-parametric Kendall's tau correlations were conducted. Similar to the parametric correlations, the nonparametric correlations highlighted a significant positive relationship between nap-associated gains in strength of size congruity (block 1) and training performance for the Malay word-pairs, $\tau = .33$, $p = .037$, but no equivalent relationship for the English word-pairs, $\tau < .001$, $p = 1.00$. The correlation implied that the poorer the participants' performance at training, the greater the nap-associated gains towards their test performance in size congruity for the new Malay words. However, the above finding should be interpreted with a degree of caution as Steiger's z-transformation indicated that there was no significant difference between the nap-associated benefit to the Malay words compared to that for the existing English words, $z = .98$, $p = .33$.

(a) (b)

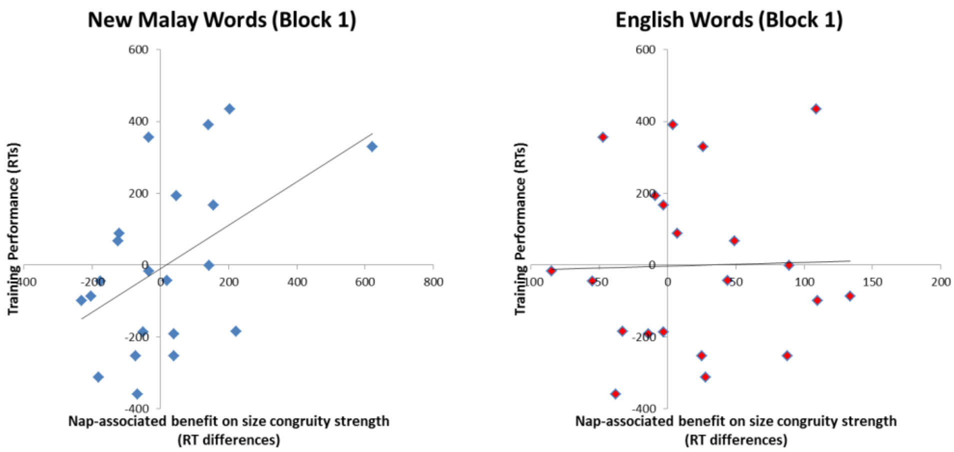


Figure 9. Experiment 2: Correlation between training performance and nap-associated gains in strength of size congruity effects in presentation block 1 trials for (a) Malay words and (b) English words

There was no main effect of session sequence [$F(1, 18) = 1.43$, $p = .25$]. However, there was an interaction between language, condition and session sequence [$F(1, 18) = 10.40$, $p = .005$] that was significant for Malay [$F(1, 18) = 13.88$, $p = .002$] but not English [$F(1, 18) = .25$, $p = .62$] trials. For the Malay words, analysis of the interaction between condition and session sequence revealed that for the wake condition, participants who completed the wake condition first had significantly quicker RTs than those who completed it after experiencing the nap condition. All other main effects and interactions for the RT analysis were not significant except the interaction between block, condition, congruity, semantic distance and the training covariate for Malay word-pairs [$F(1, 18) = 5.89$, $p = .026$]. The above finding was not expanded on as there were no theoretically relevant interactions when comparing individual variables.

Similar to Experiment 1, there was a large main effect of language (presented earlier). In contrast to Experiment 1, there was no significant interaction between language and condition [$F(1, 18) = .25$, $p = .63$]. Still, in order to investigate if there were differences in participants' variation in responses that were not reflected in mean RTs, an ANCOVA was performed on participants' CV_{RTs} with language, condition and presentation block as within-participant factors, session sequence as a between-participant factor and training performance as a covariate. There was a main effect of language [$F(1, 18) = 51.62$, $p < .001$] where CV_{RTs} for the English words were significantly smaller than CV_{RTs} for the Malay words. In contrast, there were no main effects of presentation block [$F(1, 18) = 1.01$, $p = .33$], condition [$F(1, 18) = .11$, $p = .41$] or session sequence [$F(1, 18) = .54$, $p = .47$].

An interaction of interest for the CV_{RT} analysis was a significant three-way interaction between language, condition and training performance [$F(1, 18) = 5.87$, $p = .026$]. In order to compare results with the RT analysis, the three-way interaction was further analysed in terms of difference between wake and nap $CV_{RTs.}$ A nap-associated benefit in CV_{RTs} was calculated by substracting the CV_{RTs} for the nap condition from the CV_{RTs} in the wake condition for the Malay and English words, where a greater difference would highlight a greater nap-asscoiated benefit. The nap-associated CV_{RT} benefit for the Malay and English words were then correlated with training performance. Analysis of nap-associated CV_{RT} benefits revealed that the three-way interaction was driven by the significant between training performance and nap-associated CV_{RT} benefits for the Malay words $r = -.44$, $p = .045$, where participants who had better training performance displayed greater nap-associated CV_{RT} benefits (see Figure 10a). In contrast, there was no equivalent relationship between training performance and nap-associated CV_{RT} benefits for the English words, r = .12, p = .60 (see Figure 10b). Moreover as visual examination of Figure 10a and 10b suggest that the data may deviate from normality, non-parametric Kendall's tau correlations were also conducted. The significant relationship between training performance and nap-associated CV_{RT} benefits for the Malay word-pairs remained significant when adopting the nonparametric correlation, τ = -.37, p = .023. There was no significant relationship between training performance and nap-associated CV_{RT} benefits for the English word-pairs r, τ = .068, p = .67. A Steiger's z-transformation indicated the difference between the nap-associated CV_{RT} benefit to the Malay words and that of the existing English words was not significant, z = -1.41, p = .16. Hence, the three-way interaction and correlations should be interpreted with caution.

(a) (b)

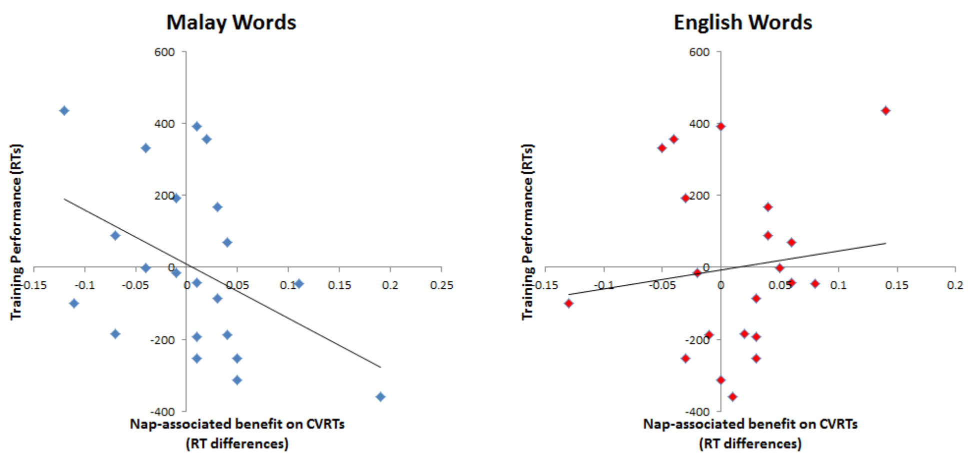


Figure 10. Experiment 2: Correlation between training performance and nap-associated benefits in CV_{RT} for (a) Malay words and (b) English words

Sleep Stage and Spindle Analysis

Table 8. Experiment 2: Sleep parameters for participants in the sleep group. Parentheses denote the standard deviation of the means.

Sleep Parameter	Mean time in minutes
Total sleep time	78 (12)
Wake time after sleep onset	8 (10)
Sleep latency	8(4)
Stage 1	8 (5)
Stage 2	33 (12)
SWS	27 (15)
REM	10 (8)

Note: SWS= slow-wave sleep, REM= Rapid eye movement sleep

Table 9. Experiment 2: Sleep spindle measures for frontal electrode. Parentheses denote the standard deviation of the means.

	F3	F4	Mean Frontal (F3 +F4)
Density	0.89 (0.55)	0.88 (0.87)	0.89 (0.48)
Count	54 (33)	52 (26)	53 (28)

Note: Total count = total count of spindles, Spindle Density = mean spindle count per minute

Similar to Experiment 1, analyses were carried out on the sleep polysomnography measures. The main sleep parameters of the participants in the nap condition are displayed in Table 8. Correlations between sleep parameters and behavioural data did not reveal any significant relationships, and were hence not expanded upon. The most common measures of reflecting spindle activity were depicted in Table 9, which was made up of the count of all NREM (stage 2 and SWS) spindles detected and NREM spindle density (mean number of spindles per minute). Similar to sleep parameters, correlations between sleep spindles and behavioural data also did not reveal any significant relationships, and were hence not elaborated.

3.2.3 Discussion

Table 10. Experiment 2: Summary of main experimental findings

Measure	Main findings

Semantic Distance Effect	• Significant main effect for both Malay and English word-pairs
	• Nap-associated gains in strength of semantic distance effect
	    o Only for Malay word-pairs
	    o Specific to participants with poor training performance

Size Congruity Effect	• Significant main effect for English word-pairs
	• No significant main effect for new Malay words
	• Nap-associated gains in strength of size congruity effect
	    o Only for Malay word-pairs in presentation block 1
	    o Specific to participants with poor training performance

CV_{RT}	• No nap-associated benefits when looking across all participants • Nap-associated CV_{RT} benefits for Malay words were displayed in participants with better training performance
Sleep Polysmonography	• No significant correlations between behavioural findings and any sleep polysomnography measures (sleep stages and sleep spindles)

A summary of the main findings from Experiment 2 can be found in Table 10. Experiment 2 expanded on the findings in Experiment 1 by investigating the effects of sleep on memory consolidation and integration of newly learnt words after a shorter duration of offline consolidation (90 minutes nap opportunity) as compared to overnight sleep. Results from Experiment 1 revealed that participants who experienced overnight consolidation displayed stronger size congruity and semantic distance effects for newly learnt words in the most salient trials, compared to participants who remained awake for an equivalent duration. Based on studies that depicted sleep-associated benefits for declarative memory occurring after a nap (Alger et al., 2012; Lahl et al., 2008; Lau et al., 2011; Takashima et al., 2006), it was hypothesized that results from Experiment 2 would be replicated those in Experiment 1.

Similar to Experiment 1, participants displayed a main semantic distance effect for both the existing English word-pairs and for the newly learnt Malay word-pairs. The above finding may be directly related to the difference between the nature of the automaticity effects previously mentioned in Experiment 1. The semantic distance effect is not affected by SRC (i.e. participants had to focus on the meanings of the words to select the larger animal). Focusing on results for the newly learnt words, after an approximately 12 hour training-test interval in Experiment 1, the semantic distance effect was present in the group that slept after training sessions and also in the group that remained awake for a comparable duration. In Experiment 2, results indicated that a 90-minute training-test interval was sufficient in eliciting semantic distance effects for the new words, regardless of whether participants slept or remained awake.

Still, in Experiment 1, overnight sleep led to sleep-associated enhancements where the sleep group displayed stronger semantic distance effects for new words (congruent trials) than the wake group. The strength of the semantic distance effect for new words in the sleep group was also directly related to SWS. Conversely, there were no equivalent sleep-associated enhancements in Experiment 2 when comparing performance across all participants for Malay word-pairs in the nap and wake sessions. Therefore, as suggested in Experiment 1, in the case where semantics are directly related to task performance and there is a limited effect of SRC, sleep may not be essential to strengthen new word knowledge. The current findings also imply that a shorter consolidation period where memories are protected from further interference may be sufficient for semantic distance effects to occur. However, the difference between Experiment 1 and 2 suggest that even if sleep may not be essential for semantic distance effects to occur, individual differences in sleep, specifically SWS (Alger et al., 2012), may lead to enhancements in semantic distance effects over wakefulness. Moreover, for these sleep-associated enhancements to occur, a longer offline consolidation period (e.g. overnight sleep) may be needed.

As mentioned in Chapter 2, the size congruity effect is thought to reflect a stronger measure of automatic semantic access which requires stronger S-R mappings to emerge, where a mismatch between physical font- referent mappings would affect participants response to select the larger (physically or semantically) item. Experiment 1 highlighted that for the newly learnt Malay words, the size congruity effect is sleep-dependent as it is only present after a period of overnight sleep (but not after an equivalent duration of wakefulness) and the strength of the effect was directly related to sleep spindle activity. In contrast, there were no size congruity effects for the newly learnt words in both nap and wake conditions in Experiment 2. One potential explanation for the difference in results between the overnight (Experiment 1) and nap (Experiment 2) studies could be due to the quantity of sleep as participants who experienced overnight sleep had, on average, more than 5 times as much TST compared to participants in the nap condition. In addition, the quality of sleep architecture such as spindle activity is also different between overnight sleep and a daytime-nap. Participants in the overnight study had more than twice the amount of NREM spindles and less variation in spindle density than those who experienced a daytime-nap. Since the size congruity effect emerges at a higher threshold, it may draw more extensively from general and global information integrated across existing knowledge. If so, the difference in size congruity effects in Experiments 1 and 2 would support findings from previous literature that both the quantity and quality of sleep are directly related to memory consolidation (Alger et al., 2012) and knowledge integration.

The difference in findings between existing nap studies (such as Takashima et al. 2006) and Experiment 2 may be due to the difference in the nature of experimental tasks used in both studies. In Takashima et al.'s (2006) study, experimenters used a recognition task where participants were tested if they remembered the picture (form) from the training session. In contrast, the paradigms used in Experiment 2 relied on processing beyond the recognition of word forms and was further linked with the semantics of the newly learnt words. In Lahl et al.'s

(2008) study, participants learnt lists consisting of existing words whereas in Experiment 2 participants had to learn both the form and meanings of new words in order to complete the task during the test session. Hence, unlike the above nap studies, sleep-dependent benefits investigated in Experiment 2 may only occur in the current experiment after a longer consolidation period. Experiment 1 also implied that overnight sleep (but not an equivalent period of wakefulness) led to enhanced automaticity in processing (CV_{RTs}) newly learnt Malay words that matched well-integrated English words. In contrast, there was no sleep-associated benefit towards CV_{RT} for Malay words in Experiment 2 after a period of daytime napping when looking across all participants. This finding further supports the suggestion that a longer consolidation period may be needed for sleep-associated gains in the processing of new words to occur.

Even though there were no clear benefits of sleep when focusing on RTs across all participants, sleep-associated effects emerged when taking into account individual differences in training performance. Due to the greater variation in training performance in Experiment 2, participants' RTs for all correct trials from both wake and nap conditions were collated and an individual training RT score was produced for each participant. Therefore between participants, those who had slower RT scores were considered to have poor training performance and those with swifter RTs were considered to have good training performance. There was a significant relationship between nap-associated gains in strength of semantic distance and training performance for the Malay word-pairs, where the poorer the participants' training performance, the greater the benefit of a nap over wakefulness on the strength of the semantic distance effect exhibited by the participants. Moreover, when examining the Malay nap-wake difference scores in Figure 9a, the nap-associated gains displayed in the x-axis seemed to only extend to participants who performed poorly during training, but not to participants with better than average training performance (i.e. negative mean-centred training RT scores in the y-axis).

A similar positive correlation between training performance and nap-associated gains was also displayed for the size congruity effect in the new Malay words. Still, the lack of a significant difference between training performance and nap-associated gains in the strength of the size congruity effect for the Malay compared to English baseline words indicates that the results should be interpreted with a degree of caution, compared to findings for the semantic distance effect where the relationship between training performance and nap-associated gains were unique to the Malay words. It should be noted that there was an opposite trend in the relationship between training performance and nap-associated benefits when looking at CV_{RT}. Participants with better training performance displayed greater nap-associated CV_{RT} benefits for the Malay words, whereas there was no equivalent relationship for the English words. This might seem contradictory to findings related to the semantic distance and size congruity effects. However, as there was no significant difference between the nap-associated CV_{RT} benefit to the Malay words compared to that for the existing English words, it is not possible to rule out the notion that circadian effects may have confounded the Malay CV_{RT} findings.

The overall findings between training performance and nap-associated benefits in Experiment 2, were supported by previous research by Drosopoulos, Schulze, Fischer and Born (2007) in their study on sleep and retroactive interference that was mentioned in Chapter 1. The study used a paired-associate training paradigm where participants initially learnt two lists (A and B) of words and each word in list A was paired with a word in list B. The level of encoding was also manipulated where the strong encoding group had more feedback and a 30% stricter training criterion than the weak encoding group. Within the strong and weak encoding groups, half the participants slept between training and testing whereas the other half remained awake. The experimenters found that following retroactive interference of new C-D paired associates (that were trained using the same method as A-B associates), performance for weakly encoded items was better after overnight sleep compared to an equivalent period of wakefulness. Moreover, there were no differences between wake and sleep groups for intensely encoded word-pairs, highlighting that the beneficiary effect of sleep against subsequent interference was unique to weakly encoded items. Similarly, if training performance was viewed as a measure of encoding level (i.e. poorer training performance would be linked to weaker encoding of new items), then the current experiment would concur with the above finding in overnight sleep to a large extent as the benefits of a nap only occurred for the participants who performed poorly but not strongly during training.

Combining Experiment 1 and 2 findings on the semantic distance and size congruity effects in newly learnt words, additional insights into the CLS model (McClelland et al., 1995) and CLS model of word learning (Davis & Gaskell, 2009) can be gained especially in terms of the time course of memory consolidation. In the original CLS model, researchers proposed that the transfer of memories from the hippocampal to neocortical system would require an extended period of time (e.g.

days or years). In contrast, word learning studies have found measures of lexical integration after the course of an overnight sleep (Davis et al., 2009; Dumay & Gaskell, 2007; Tamminen et al., 2010). Experiment 1 supports previous word learning studies highlighting that measures of hippocampal to neocortical transfer and integration can occur after an overnight sleep. Experiment 2 extends this finding by suggesting that there may be a minimum duration of offline consolidation required for integration of new memories into existing neocortical networks. Although there has been no existing research denoting a minimum duration of sleep for memory consolidation to occur, results suggest that a 90-minute period of offline consolidation may not be enough for sufficient hippocampal to neocortical transfer to display sleep-dependent size congruity effects in newly learnt words. This could be due to the lack of quantity of sleep or the poorer quality of sleep in a 90-minute nap. Findings suggest that both the quantity and quality of sleep architecture may be essential, where a minimum duration of SWS coupled by sufficient quality of sleep spindle activity is needed to provide the ideal platform for the reactivated memories during SWS to be transferred from the hippocampal to neocortical system. (For further discussion on CLS please refer to General Discussion in section 3.4).

In addition, comparing the overnight sleep data in Experiment 1 with the nap data in Experiment 2 also adds new insights on how the duration of sleep may affect the nature in which sleep-dependent benefits in memory consolidation are manifested. It is possible that for shorter sleep durations such as daytime naps, sleep associated benefits in memory consolidation may be selective to differences in quality of learning (or training). For example, sleep would be predicted to be more crucial towards participants who are poorer learners (Experiment 2). In contrast, a longer duration of sleep seems to benefit memory consolidation across all learners as compared to an equivalent period of wakefulness. For longer sleep durations, individual differences in behavioural performance also seem related to specific sleep parameters such as SWS and sleep spindles. However, neither Experiment 1 nor Experiment 2 have directly examined if within participants, sleep also selectively benefits automaticity of processing new words that are less well encoded over well encoded words. None of the systems consolidation models have specifically outlined how general individual differences in learning may interact with sleep and what implications that would have for the memory transfer from the hippocampal to neocortical system. Further investigation that manipulates levels of training or encoding within individuals would be valuable in gaining more understanding regarding the conditions that promote or relate to sleep-dependent memory consolidation.

3.3 Experiment 3

One of the key findings in Experiment 2 was that nap-associated gains in strength of semantic distance and size congruity effect for Malay words where specific to participants who performed poorly during training. As mentioned in the discussion for Experiment 2, the above finding was supported by Drosopoulos et al., (2007) who found that following retroactive interference from a new set of word-pair associates, performance on initial weakly encoded word-pairs was better after sleep compared to an equivalent period of wakefulness. Moreover, participants in Drosopoulos et al.'s (2007) study did not show any sleep-dependent gains for items that were strongly encoded. Still, both Experiment 2 and Drosopoulos et al.'s (2007) study examined encoding difficulty by considering between-participant differences in training performance. It is not possible to conclude from either study if sleep also selectively benefits automaticity of processing new words that are less well-encoded over words that are well-encoded when the encoding is manipulated within participants.

There has been limited work where strength of encoding was manipulated as a within-subjects factor. One study that manipulated levels of encoding within-subjects was a study conducted by Schmidt et al. (2006). The participants in the study learnt two sets of word-pair associates on separate occasions where one set of words had more concrete associations and the other set had less concrete associations. Based on participants' immediate recall performance, the experimenters termed the more concrete set of word-pair associates as 'easy-encoding' and the less concrete word-pair associates as 'difficult-encoding'. After a four-hour nap opportunity, the experimenters found no difference in improvements relating to sleep-dependent consolidation between 'easy-encoding' and 'difficult-encoding' word-pairs. Still, even though encoding was manipulated as a within-subjects factor in Schmidt et al.'s (2006) study, participants experienced easy and difficult encoding conditions on separate occassions. Therefore, it was not possible in Schmidt et al.'s (2006) design to directly investigate if sleep selectively benefits the consolidation of weakly encoded (difficult-encoding) items over strongly encoded (easy-encoding) items. There was also no condition where participants remained awake after learning the word-pair associates. Hence, it was not possible to investigate any potential benefits of napping over wakefulness.

Therefore, Experiment 3 aimed to investigate if the nap-associated findings between training performance and automaticity in Experiment 2 would also be displayed when the level of encoding was manipulated within-participants. Unlike Schmidt et al.'s (2006) study, Experiment 3 examined if sleep would selectively benefit weakly encoded items over strongly encoded items and also if the benefits of sleep-associated consolidation would be significantly greater than that of an equivalent period of wakefulness. In Experiment 3, variation in encoding was manipulated within-participants whereby for each participant in the same training session, half the new words were exposed 8 times during training (weak encoding) and the other half exposed for 64 times (strong encoding). Even though Schmidt et al. (2006) did not report any differences between levels of encoding and subsequent sleep dependent consolidation, as mentioned previously, the experimental design did not allow for comparisons of sleep-associated benefits in consolidation over wakefulness and the manipulation of encoding was done on separate sessions. Based on findings from Experiment 2 and Drosopoulos et al.'s (2007) study, it was predicted that if between-subjects nap-associated gains would parallel within-subject benefits, then the new words that were weakly encoded would experience greater sleep associated gains over wakefulness, in the size congruity and semantic distance paradigms compared to new words that were strongly encoded.

In addition, using sleep polysomnography, Experiment 3 also aimed to investigate if specific sleep architecture would be related to the above within-participant variations. It is important to note that there were no relationships between sleep architecture and between participant differences in behavioural findings in Experiment 2. Similar to the behavioural findings, existing research on the relationship between individual sleep components, variation in encoding and post-sleep performance is also not conclusive. Drosopoulos et al., (2007) did not report any relationships between sleep parameters and weak or intense encoding for the paired-associate tasks, while another study by Tucker and Fishbein (2008) found that SWS was positively correlated to post-training improvements in word-pair recall for participants who performed well but not poorly during learning. The above relationship mirrored that of a previous word learning study which found a positive correlation between SWS and memory strength for new words (Tamminen et al., 2010). Due to the inconclusive nature of previous research, there were no specific predictions related to sleep polysomnography data. Still, out of all sleep architecture, it was most plausible (based on current literature) that SWS would play a role in moderating the within-participant effects of encoding on subsequent behavioural performance.

3.3.1 Methods

Apparatus and Materials

The test apparatus used was the same as Experiment 2. The experimental stimuli for the whole experiment included 24 Malay words with 24 English ‘translations’ and a further 24 existing English animal names (baseline) adapted from Experiment 2, where participants saw half the stimuli in the wake condition and the other half in the nap condition. The Malay words consisted of the 18 words from Experiment 2 and 6 additional words from the same database. The English animal names consisted of the 36 animal names in Experiment 2 and an additional 12 animals from the database used in Experiment 2. The selected animal names were ranked according to mean size norms and allocated to 3 size groups: Small (M = 1.20, SD = 0.35), Medium (M = 2.61, SD = 0.44) and Large (M = 5.01, SD = 0.58), with 16 animals in each group. Items in each size group were matched whereby an ANOVA revealed no significant differences in lexical decision RTs, frequency ratings and word length between groups (all p’s >.05).

Two separate sets of Malay-English (animal name) pairs were created. For each set, 8 small, 8 medium and 8 large English animal names were selected from the size ranking databases presented in Experiment 2 and a Malay word was randomly paired with each English animal name. Next, half the items from every size ranking group were randomly allocated to the ‘weak encoding’ condition and the other half to the ‘strong encoding’ condition. Finally, the items were distributed into 2 groups (A or B) for the wake and nap condition (see Appendix B for full list of items). A more comprehensive overview of the materials can be found in section 3.2.1 (Apparatus and Materials).

Stimuli and Design

Individual trials in the training and test sessions were manipulated in the same way as Experiment 2. However, due to the within-participants manipulation of encoding strength, the design within each session varied from Experiments 1 and 2. The number of trials was also different as participants learnt 12 new items instead of 9 in Experiment 2.

For the 2-AFC feedback task in the training session, 6 items were allocated to the 'weak encoding' condition and the remaining 6 items were allocated to the 'strong encoding' condition (see Appendix 5). Each of the weakly encoded items was viewed for a total of 8 presentations – 4 times as correct choices to the target and 4 times as foils. In contrast, each strongly encoded item was presented 64 times - 32 times as correct target choices and 32 times as foils. Target items were always paired with foils of the same strength and were counterbalanced on the left and right to prevent response biasness. Therefore, participants were presented with 216 (2-AFC) trials in the training session, consisting of 24 weakly encoded trials (6 items x 4 targets) and 192 strongly encoded trials (6 items x 32 targets).

For the test session, items differed in relative semantic distance and size congruity similar to Experiment 2. In order to have a similar number of trials as Experiment 2, it was not possible to generate all possible item-pair combinations. Instead, for the Malay comparisons, participants saw 32 item-pairs with small semantic distances (16 small-medium animals and 16 medium-large animals) and 32 item-pairs with large semantic distances (32 small-large animals). Out of the 64 unique pairings, each newly learnt Malay word was viewed a total of 8 times. In addition, within small-medium, medium-large and small-large pairs, the strength of the target (correct choice) and foil was also manipulated for each trial where word-pairs could consist of one of four target-foil pairs: strong(target)-strong(foil), strong-weak, weak-strong, weak-weak, where the target-foil strength was counterbalanced. In summary, participants saw 256 trials for the Malay comparisons, 64 target-foil counterbalanced word-pairs (32 small and 32 large semantic size distance-pairs) x 2 (congruity) x 2 (left/right counterbalancing). Participants also saw 256 trials for the English comparisons. The English trials were manipulated in a similar way to the Malay trials with the exception of the target-foil strength manipulation. Within the Malay and English comparisons, the trials were also allocated to 4 separate counterbalanced blocks.

A mixed-subjects design was used for the Malay comparisons. The dependent variable was RTs for correct responses. There were six within-subjects independent variables, namely: experimental condition (wake/nap), congruity (congruent/incongruent), semantic size distance (small/large), target strength (strong/weak), foil strength (strong/weak) and block of presentation (first/second/third/forth presentation). The between subject variable was session sequence (wake condition first or nap condition first). For the English comparisons, a five way mixed-subjects design was used, with all the above factors excluding target and foil strength. Separate analyses were also conducted with proportion of errors as a dependent variable (presented in Appendix 1.3).

A measure of general sleep quality was taken using the PSQI. In addition to the SSS, participants in Experiment 2 also completed a 2-AFC RT alertness task, which consisted of 18 randomized trials where participants saw either a '0' or '1' on the screen and had to make the relevant response as quickly and accurately as possible. This task was previously used in sleep studies (van der Helm, Gujar, Nishida, & Walker, 2011) as a more objective measure in addition to the SSS which was deemed as a more subjective measure. A paper-based matching task was also generated for each group of items within each stimuli set. The matching task consisted of two columns of words; the right column consisted of the newly learnt Malay words and the left column consisted of the English words the Malay words were paired with. The positions of both the Malay and English words were randomised.

Participants

Participants were 22 monolingual native English speakers (10 males, 12 females; *M* = 20.5, range: 18-23). Participants were randomly allocated to the wake first (n = 10) or nap first (n = 12) experimental sequence. All remaining recruitment procedures were replicated from Experiment 2.

Procedure

The procedure for Experiment 3 was similar to Experiment 2, where each participant underwent 2 training-test sessions both carried out in the University of York Psychology departmental sleep laboratory. The main difference between both experiments was the number of trials participants viewed in the training session and test sessions.

Training. Similar to previous experiments, participants were first shown exposure trials during the training session. There were 12 different trials representing the 12 Malay words in the stimuli set. However unlike previous experiments were the stimuli were repeated in 3 blocks, participants only saw each exposure trial once in Experiment 3. The procedure for the 2-AFC feedback trials was the same as Experiment 2.

Test. Apart from difference in the number of trials viewed, participants in Experiment 3 also completed 4 blocks of English-Malay comparison tasks (see Stimuli and Design section) instead of the 2 blocks completed by participants in Experiment 2.The remaining procedure within the test session was identical to Experiment 2.

After each training and test session, participants also completed the relevant matching task related to the set of stimuli they were allocated to. For the matching task, participants were instructed to match the Malay words on the right column to their trained English meaning on the left column. At the end of whole experiment, the vocabulary and matrix reasoning subtests from the WASI (Wechsler, 1999) were administered. Due to time constraints, the WRAT-3 test was not administered as it was not related to any behavioural or sleep measure in Experiment 2.

Polysomnographic (PSG) recording and EEG Analysis. The PSG setup and EEG analysis in Experiment 3 was identical to that in Experiment 2.

3.3.2 Results

Data from 2 participants who did the wake first experimental sequence were excluded: one participant had WASI scores more than 3 SDs from the overall participant mean and the other participant was excluded due to equipment fault. There were no significant differences in SSS ratings and accuracy levels in RT alertness task between training and testing in both wake and nap conditions for the remaining 20 participants (all p's $> .28$). For the paper-based matching task, participants performed significantly worse after the test session as compared to the training session [$F(1, 19) = 9.61$, $p < .006$], but performance was not affected by the condition participants experienced [$F(1, 19) = .60$, $p = .80$].

Training Session

Table 11. Experiment 3. Mean response time (in milliseconds) and error rates (percentage) for the training session. Standard deviations are presented in parentheses.

	Block 1				**Block 2**			
	Wake Condition		Nap Condition		Wake Condition		Nap Condition	
Encoding	Weak	Strong	Weak	Strong	Weak	Strong	Weak	Strong
RT(ms)	1353 (341)	1124 (346)	1343 (340)	1182 (346)	1169 (348)	1064 (334)	1213 (317)	1104 (331)
Error Rate (%)	16.2 (14.9)	6.4 (5.2)	14.9 (17.5)	5.6 (5.3)	8.6 (12.8)	4.3 (3.2)	8.6 (8.8)	4.9 (3.6)

Performance in the training session was analysed in 2 separate blocks (separated by the untimed break) and presented in Table 11. The mean error rate for weak (12.1%) encoding items was more than twice of strong (5.3%) encoding items. This suggests that the manipulation in number of presentations during training was effective in affecting level of encoding.

Test Session

RTs went through the same trimming procedure as Experiment 2. As there were 4 experimental blocks in Experiment 3, items from blocks 1 and 2 were grouped into one block and items from blocks 3 and 4 were grouped into another block for data trimming in order to reduce average data loss. This resulted in an average data loss of 4.9% and 4.3% for Malay and English comparisons respectively. For the Malay comparison, a seven-way mixed ANOVA was performed with session sequence (wake condition first or nap condition first) as the between-participant factor, and condition (wake/nap), presentation block (1, 2, 3 or 4), target strength (weak/strong) foil strength (weak/strong), congruity and semantic distance as within-participant factors. Interactions with presentation block and session sequence were not reported except for cases of theoretical interest.

As the encoding strength was only manipulated for newly learnt Malay words, a five-way ANOVA was performed for English comparisons with all the above factors except target strength and foil strength. It should be noted that since the number of observations (single cells) are different for the Malay and English comparisons, the degrees of freedom (df) reported in the results section will differ between the Malay and English items. In addition, the df2 for a within-subjects design (and mixed subjects design) is calculated using a different method than that for a between-subjects design[11].

Table 12. Experiment 3: Mean response time (in milliseconds) for all correct responses after data trimming. Standard errors are presented in parentheses.

	English		Malay	
	Wake	**Nap**	**Wake**	**Nap**
Block 1				
Congruity				
Congruent	804 (32)	861 (26)	1491 (111)	1504 (155)
Incongruent	819 (34)	861 (27)	1516 (114)	1524 (93)
Semantic Distance				
Small	856 (35)	916 (31)	1567 (109)	1606 (106)
Large	766 (31)	806 (21)	1440 (115)	1422 (101)
Target				
Strong			1440 (106)	1464 (99)
Weak			1567 (124)	1564 (110)
Foil				
Strong			1527 (114)	1489 (91)
Weak			1479 (111)	1539 (116)
Block 2				
Congruity				
Congruent	774 (36)	827 (32)	1273 (103)	1318 (104)
Incongruent	774 (29)	828 (32)	1351 (111)	1317 (89)

[11] A detailed method of calculating df2 for within-subjects design can be found at: http://egret.psychol.cam.ac.uk/psychology/graduate/Guide_to_ANOVA.pdf

Semantic Distance				
Small	833 (37)	877 (35)	1386 (109)	1384 (102)
Large	715 (28)	778 (30)	1238 (105)	1252 (93)
Target				
Strong			1298 (98)	1318 (96)
Weak			1327 (119)	1318 (100)
Foil				
Strong			1364 (117)	1331 (94)
Weak			1261 (97)	1305 (100)
Block 3				
Congruity				
Congruent	778 (35)	809 (40)	1158 (98)	1201 (101)
Incongruent	788 (40)	804 (36)	1184 (97)	1173 (82)
Semantic Distance				
Small	827 (40)	857 (40)	1241 (107)	1278 (101)
Large	739 (36)	755 (36)	1101 (88)	1095 (78)
Target				
Strong			1162 (88)	1187 (93)
Weak			1180 (109)	1186 (85)
Foil				
Strong			1167 (101)	1195 (93)
Weak			1175 (94)	1178 (85)
Block 4				
Congruity				
Congruent	758 (38)	790 (35)	1086 (82)	1114 (88)
Incongruent	771 (38)	784 (30)	1197 (110)	1130 (91)
Semantic Distance				
Small	812 (41)	841 (35)	1199 (95)	1201 (93)
Large	717 (34)	733 (29)	1084 (96)	1043 (85)

Target		1134 (86)	1130 (92)
	Strong	1149 (107)	1114 (84)
	Weak		
Foil		1164 (101)	1132 (95)
	Strong	1119 (91)	1112 (84)
	Weak		

Note: Strength of Target and Foil was not manipulated for the English comparisons

Response time measures are summarised in Table 12. The semantic distance effect was significant in both the English [$F(1, 17) = 148.68, < .001$] and Malay [$F(1, 13) = 88.14$, $p < .001$] comparisons. In contrast, there was no size congruity effect for the English comparisons [$F(1, 17) = .07$, $p = .80$] and marginally significant size congruity effect for the Malay comparisons [$F(1, 13) = 4.58$, $p = .052$]. There was a main effect of presentation block for the English [$F(3, 51) = 6.70$, $p = .001$] and Malay [$F(3, 39) = 96.51$, $p < .001$] comparisons where participants had faster RTs with increased practice. There was also a main effect of condition for the English [$F(1, 17) = 6.26$, $p = .023$] comparisons where participants had overall faster RTs when they remained awake as compared to the test session after a nap. This suggests that potential fatigue effects may be present in the nap condition that may be driven by poor quality sleep or sleep inertia. There was no main effect of condition for the Malay comparisons [$F(1, 13) = .007$, $p = .93$].

When looking at the encoding strength manipulations for the Malay comparisons, the main effect of target strength was not significant [$F(1, 13) = 2.42$, $p = .14$]. However, there was a significant interaction between block and target strength [$F(3, 39) = 5.53$, $p = .003$] where participants had faster responses for strong compared to weak targets in the first presentation block. The effects of target strength seemed to diminish with practice and were not significant in the other presentation blocks. The main effect of foil strength was marginal [$F(1, 13) = 4.35$, $p = .057$] and results indicated that in general, strong foils slowed down participants processing of the newly learnt items to a greater extent than weak foils.

The main hypothesis was that new words that were weakly encoded would experience greater sleep associated gains in the size congruity and semantic distance paradigms compared to new words that were strongly encoded. Therefore, even though there were no significant interactions between encoding strength (for targets and foils) and the size congruity or semantic distance effect, further analysis was conducted to investigate if interactions with size congruity or semantic distance would be displayed when considering separate combinations of weakly encoded and strongly encoded target-foil pairs. All possible target-foil pairings for the strongly and weakly encoded Malay words displayed significant semantic distance effects (all p's <.001)[12]. In contrast, the size congruity effect was only present in Malay word-pairs where both target and foils were strongly encoded during training [$F(1, 17) = 5.07$, $p = .038$], but not when the strongly encoded targets were paired with weak foils [$F(1, 17) = .74$, $p = .40$]. There was also no size congruity effect for weakly encoded targets regardless of whether they were paired with strong [$F(1, 17) = .72$, $p > .41$] or weak [$F(1, 17) = 1.58$, $p = .23$] foils.

Still, in order to compare findings with Experiment 2, further difference scores for each ANOVA variable was calculated by subtracting RTs in the nap session from RTs in the wake session for the Malay words. Similar to Experiment 2, these difference scores were predicted to highlight the potential benefits of nap over wakefulness for memory consolidation of the Malay words. Analysis of the difference scores revealed a marginal interaction between foil strength and semantic distance [$F(1, 18) = 4.37$, $p = .051$], where there was a significantly larger nap-associated benefit for the strength of the semantic distance effect in Malay word-pairs with strongly encoded foils (see Figure 11) compared with word-pairs with weakly encoded foils. In addition, when separating word-pairs with strongly encoded and weakly encoded foils, only the Malay word-pairs with strongly encoded foils displayed nap-associated gains in the semantic distance strength [$F(1, 18) = 5.77$, $p = .027$]. There were no nap-associated gains in the semantic distance strength for the Malay-word pairs with weakly encoded foils [$F(1, 18) = .16$, $p = .90$].

[12] Semantic distance effect for

- Strong target and strong foil: [$F(1, 17) = 49.67$, $p < .001$]
- Strong target and weak foil: [$F(1, 17) = 31.03$, $p < .001$]
- Weak target and strong foil: [$F(1, 17) = 34.75$, $p < .001$]
- Weak target and weak foil: [$F(1, 17) = 39.49$, $p < .001$]

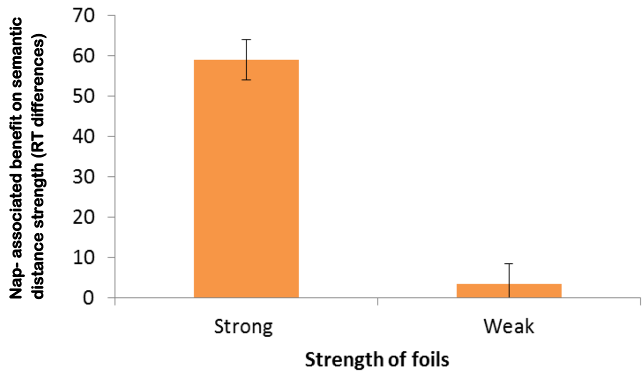


Figure 11. Experiment 3: Effect of foil strength on nap-associated benefit for semantic distance effect

The difference scores analysis and revealed a significant three-way interaction between target strength, congruity and distance [$F(1, 18) = 6.46$, p = .20]. Further analysis highlighted that the above relationship was driven by the significant interaction between congruity and distance for Malay word-pairs with weakly encoded targets [$F(1, 18) = 6.54$, p = .20]. Figure 12 depicts that for weakly encoded targets, there was a larger nap-associated benefit for size congruity effect for Malay word-pairs with large semantic distances compared to small semantic distances. In contrast, there were no equivalent nap-associated benefits for the word-pairs with strongly encoded targets [$F(1, 18) = .23$, p = .64]. It should also be noted that the other two possible interactions stemming from the target strength, congruity and distance interaction did not reach significance (all p's > .08). Further analysis of the Malay word-pairs in Figure 12 revealed that word-pairs with strongly encoded targets and larger semantic distances displayed a marginally significant nap-associated benefit on size congruity effect [$F(1, 18) = 4.04$, p = .06]. There were no equivalent significant or marginally significant nap-associated benefits on size congruity effect for the other combinations of word-pairs[13].

[13] For Malay word-pairs with weak targets and small semantic distances: [F(1, 18) = .72, p = .41]
For Malay word-pairs with strong targets and small semantic distances: [F(1, 18) = .1.51, p = .23]
For Malay word-pairs with strong targets and large semantic distances: [F(1, 18) = .1.68, p = .21]

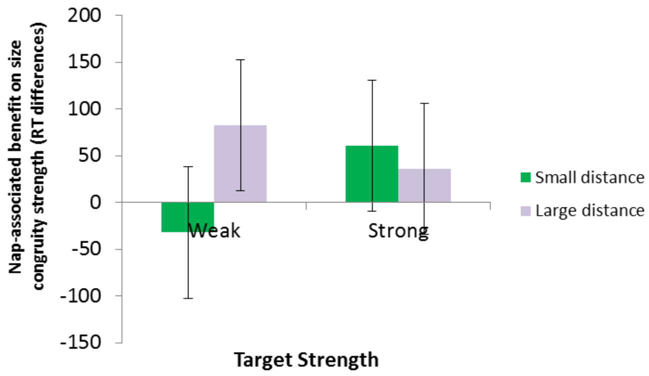


Figure 12. Experiment 3: Effect of target strength and semantic distance on nap-associated benefit for size congruity effect

Difference scores between nap and wake sessions were also calculated for the English words, but ANOVAs did not reveal any significant findings (all p's > .45). All other main effects and interactions (for both the RT and difference scores analysis) were not significant except the interaction between congruity and distance in the English comparisons [$F(1, 17) = 4.91$, $p = .041$]. The finding was not expanded as there were no theoretically relevant main effects when looking at individual factors.

Similar to Experiment 2, in order to investigate if there were differences in participants' variation in responses that were not reflected in mean RTs, a three-way mixed ANOVA was performed on participants' CV_{RTs} with language, condition and presentation block as within-participant factors and session sequence as a between-participant factor. There was a significant main effect of language [$F(1, 18) = 48.97$, $p < .001$] where participants had significantly larger CV_{RTs} for Malay compared to English words. There was also a significant main effect of presentation block [$F(1, 18) = 8.63$, $p < .001$] where CV_{RTs} increased with each presentation block, which may be due to a potential fatigue effect within participants. There were no main effects of condition [$F(1, 18) = .16$, $p = .69$] or session sequence [$F(1, 18) = 1.09$, $p = .31$] and no theoretically relevant interactions (all p's > .47).

Sleep Stage analysis

Table 13. Experiment 3: Sleep parameters for participants in the sleep group. Standard deviations of the means are in parentheses.

Sleep Parameter	Mean time in minutes
Total sleep time	60 (22)
Wake time after sleep onset	18 (19)
Sleep latency	17(8)
Stage 1	14 (7)
Stage 2	27 (13)
SWS	14 (15)
REM	5 (7)

Note: SWS= slow-wave sleep, REM= Rapid eye movement sleep

The main sleep parameters of the participants in the nap condition are displayed in Table 13. Similar to Experiment 2, correlations between sleep parameters and behavioural data did not reveal any significant relationships, and hence were not expanded upon. In Experiment 1, sleep spindle activity was uniquely related to the strength of the size congruity in new words. As there was no main effect of size congruity for the Malay words, spindle analysis was not conducted for Experiment 3.

3.3.3 Discussion

Table 14. Experiment 3: Summary of main experimental findings

Measure	Main findings
Strongly encoded Malay words (as targets)	• Significant semantic distance effects • Significant size congruity effect only when words were paired with strongly encoded foils
Weakly encoded Malay words (as targets)	• Significant semantic distance effects • No size congruity effects
Semantic Distance Effect	• Significant main effect for both Malay and English word-pairs • Nap-associated gains in strength of semantic distance effect o Only for Malay word-pairs with strongly encoded foils
Size Congruity Effect	• No significant main effect for both Malay and English words • Nap-associated gains in strength of size congruity effect o Marginal effects for Malay word-pairs with weakly encoded targets and large semantic distances
CV_{RTs}	• No nap-associated benefits
Sleep Polysmonography	• No significant correlations between behavioural findings and any sleep polysomnography measures

Experiment 3 investigated if within-participant variations in encoding strength would lead to different sleep-associated gains in size congruity and semantic distance measures of knowledge integration for new words as compared to the between-participant differences in Experiment 2. It was predicted that the new words that were weakly encoded would experience greater sleep-associated gains in the size congruity and semantic distance paradigms as compared to new words that were strongly encoded. The main findings for Experiment 3 are summarized in Table 14.

Regardless of whether they were experiencing the wake or nap condition, participants displayed a significant semantic distance effect for both the new Malay words and existing English words. This result reinforces the claim made in Experiment 2 where a 90 minute training-test interval was sufficient in producing semantic distance automaticity effects for the new words. Moreover, for the Malay words, the semantic distance effect was present for both the strongly encoded and weakly encoded words. Therefore in addition to above implications regarding the time-course of semantic distance effects to develop for second language learning, Experiment 3 also suggests that semantic distance automaticity effect can be present after limited training or weak encoding for new second language words. This is because that unlike Experiment 1 and 2 where each new Malay word was viewed 24 times in the 2-AFC training phase, each weakly encoded Malay words in Experiment 3 was only presented 8 times. Still, despite having fewer presentations than the words in Experiment 1 and 2, the weakly encoded Malay words in Experiment 3 stimulated significant semantic distance effects.

As for the size congruity effect, no main effect was present across both the English and Malay word-pairs. However, when separating the Malay word-pairs into all possible target-foil combinations based on encoding strength, there was a significant size congruity effect for Malay word-pairs where both the targets and foils were strongly encoded. There was no size congruity effect for the Malay word-pairs when strong targets were paired with weak foils. Therefore, encoding strength is important in aiding the processing of targets and foils. As mentioned in the review in Chapter 1, saliency is an important modulating factor for memory consolidation. Van Dongen et al. (2012) trained participants in a picture-location paired-associate task where the saliency item-pairs was manipulated in terms of future relevance. The experimenters told participants that out of the two sets of paired-associates that they learnt, only one set of paired-associates would be tested. Regardless of whether participants slept or remained awake between training and testing, participants remember more picture-location item-pairs from the set of paired-associates that they were told would be tested compared to the set of pair-associates that they thought would not be tested. Therefore, overall memory performance for the item-pairs that were considered to have greater saliency and future relevance was better than performance for items that were less salient. It is possible that due to the increased exposure during training, the strongly encoded words in Experiment 3 were deemed more salient and more relevant to future testing than the weakly encoded words and hence only the strongly encoded Malay word-pairs exhibited significant size congruity effects.

One of the key findings for Experiment 3 was also that nap-associated benefits in the semantic distance and size congruity effects for Malay words only emerged when taking into account the variations in encoding strength of the new words. Participants displayed significant nap-associated semantic distance effect benefits for Malay word-pairs with strongly encoded foils regardless of target strength (there was no three-way interaction between foil, semantic distance and target strength). There was also a marginal nap-associated benefit in size congruity effect that was specific towards Malay word-pairs with weakly encoded targets and large semantic distances. For the overall ANOVA, there was a marginal effect of foil strength indicating that strong foils slowed down participants processing of the newly learnt items to a greater extent than weak foils. Participants also had significantly slower RTs for Malay word-pairs with weakly encoded targets compared to strongly encoded targets for the first presentation block. Therefore, instead of the experimental prediction that greater sleep-associated gains in the size congruity and semantic distance paradigms would emerge for new words that are weakly encoded, the current findings suggest that greater sleep-associated gains in the size congruity and semantic distance paradigms were presented for Malay word-pairs that were considered to be more ‘difficult’ to process. An area for future research would be to dissociate encoding strength from processing difficulty. It seems that weakly encoded items may not always be difficult to process. Since nap-associated gains emerged for new items that were difficult to process, it may be useful to manipulate encoding difficulty instead of encoding strength in future studies. Instead of varying the number of presentations during training (encoding strength), future experiments can manipulate the quality of the presentations to increase or decrease ease of encoding – for example, when teaching half of the new words during exposure trials, the existing word could be excluded (just showing an image of the item) to make the new words more difficult to encode.

It was predicted that SWS would play a role in moderating the within-participant effects of encoding on subsequent behavioural performance. This is because in Tucker and Fishbein’s (2008) study, SWS (as a percentage of TST) was positively correlated to post-training improvements in word-pair recall for participants who performed well, but not poorly, during learning. In contrast, in

Experiment 3, there were no significant correlations between behavioural findings and any sleep polysomnography measures including SWS. One of the reasons for the lack of relationships between sleep parameters and behavioural findings could be the large variability in nap quantity and quality. Participants in Experiment 3 had approximately 14 minutes of SWS, which accounted for 23% of TST. In contrast, 21 minutes of mean SWS in Tucker and Fishbein's (2008) study accounted for 43% of TST, which was about twice of that in Experiment 3. Therefore as mentioned in both Experiment 1 and 2, a minimum duration and quality of SWS may be needed to enable active sleep-dependent offline consolidation.

A final point to note was that there were limitations in Experiment 3 that were not present in Experiment 1 or 2. In both Experiment 1 and 2, there were significant or marginally significant semantic distance and size congruity effects for the existing English word-pairs. However, in Experiment 3, the size congruity effect for the English word-pairs was not significant (or marginally significant). The lack of size congruity effects for the English word-pairs may be related to potential fatigue effects across participants. Moreover, participants had faster RTs for the English word-pairs in the wake condition compared to the nap condition in Experiment 3. The above main effect of condition was not present for the English word-pairs in Experiment 1 and 2 and suggests potential fatigue effects in the nap condition in Experiment 3 that may be driven by poor quality sleep or sleep inertia.

3.4 Chapter Summary

Experiment 2 highlighted sleep-associated benefits for the size congruity and semantic distance effects also interacted with between-participant training factors, whereby the nap-associated gains in the automaticity effects occurred only for participants who performed poorly during training (had weak encoding of the stimuli during training). In Experiment 3 variation in encoding/training was manipulated within-participants, where half of the new Malay words were exposed 8 times during training (weak encoding) and the other half exposed 64 times (strong encoding). As discussed in Experiment 3, instead of the experimental prediction that greater sleep-associated gains in the size congruity and semantic distance paradigms would emerge for new words that are weakly encoded, the current findings suggest that greater sleep-associated gains in the size congruity and semantic distance paradigms were present for Malay word-pairs that were considered to be more 'difficult' to process.

Still, it is important to note that Experiment 2 examined variations in encoding that were specific to individual differences in the general ability to learn new words, whereas Experiment 3 investigated strength of encoding that was stimuli-specific. In order to gain insights on the effects of sleep on variations in training and encoding in word learning, it may be useful to compare Experiments 2 and 3 to a further study where both participant-specific (individual differences) and stimuli-specific factors are manipulated; for example a study comparing good and poor "encoders/learners" who are trained on both strongly and weakly encoded new words. Current findings may not be contradictory; rather, it is plausible that for participant-specific differences, poor encoders/learners experience greater sleep-associated benefits than participants who are able to learn the new words very well in the training session. However, for stimuli-specific differences, ease (or difficulty) of processing the new words may play a greater role in modulating sleep-associated gains in automaticity.

Most existing research on selective factors that modulate sleep-dependent memory consolidation have focused on stimuli specific manipulations (Drosopoulos et al., 2007; Takashima et al., 2006) rather than differences between participants (Tucker & Fishbein, 2008). The above studies either have shown conflicting findings or would need further replication. Still, in most cases, studies have suggested that participant- and stimuli-specific encoding variations affect the quality of sleep-dependent memory consolidation. Stickgold (2009) has suggested that sleep may not have a beneficial effect for very strongly and very weakly encoded factors, but that a benefit of sleep would emerge for items that are encoded to a moderate level. This is because items that are very strongly encoded may not need further sleep-dependent consolidation and very weakly encoded items may be forgotten prior to sleep, whereas moderately encoded items leave room for sleep-dependent gains to emerge. Stickgold's (2009) hypothesis seems aligned with the different trends of results in existing research. It also provides further insights to the limited sleep-associated effects in Experiment 3, where it is possible the experimental manipulation in encoding strength may have been too extreme (strong/weak). It would be useful in further replications to have different levels of strong and weak encoding to see if Stickgold's (2009) hypothesis fits with the new word learning data.

As mentioned in section 3.2, one weakness with a nap design is the large variability in sleep quality. Therefore when investigating the active role of individual sleep parameters on knowledge integration in new word learning, a nap design may be a less sensitive measure compared to overnight sleep. Comparisons between Experiment 1 (overnight study) and Experiment 2 and 3 (nap studies) supports previous literature (Alger et al., 2012; Lahl et al., 2008; Schabus et al., 2005), suggesting that both the quantity and quality of sleep architecture were essential to declarative memory consolidation. The lack of sleep-associated differences in the overall size congruity and semantic distance effects in Experiment 3 (as compared to Experiment 2) may be due to the poorer quality of sleep experienced by participants. In addition to having a lower mean total sleep period, participants in Experiment 3 had more than twice the amount of sleep latency and wake time after sleep onset than participants in Experiment 2. Differences in Experiments 1 and 2 also suggested that a minimum duration of SWS coupled with sufficient quality of sleep spindle activity may be needed for knowledge integration. Participants in Experiment 2 had approximately twice as much SWS than participants in Experiment 3. Hence, it is not surprising that participants in Experiment 3 did not show greater size congruity and semantic distance in the nap compared to wake condition.

Still, differences between the overnight and nap experiments may be useful in examining if the duration of TST would predict changes in the role of sleep on memory consolidation of new words. Experiment 1 supports an active role of sleep on knowledge integration of newly learnt words as SWS and sleep spindle activity were respectively uniquely correlated to benefits in the strength of semantic distance and size congruity effects. In contrast there were no relationships between sleep polysomnography measures and behavioural findings for the new Malay word-pairs in both Experiment 2 and 3. One interpretation is that sleep only plays an active role in knowledge integration when certain conditions are present, such as longer TST and possible greater SWS (see previous paragraph). With a shorter time for memory consolidation (nap), sleep seems to play a more passive role towards knowledge integration, where nap-associated gains were still present but there were no active relationships between automaticity effects of individual sleep parameters.

In summary, both Experiment 2 and 3 have expanded on the notion mentioned in the discussion section of Experiment 1 (see Chapter 2 section 2.2.3); that for newly learnt words, the size congruity effects may require different amounts of consolidation compared to the semantic distance effects. Participants from all three experiments displayed semantic distance effects for the new Malay words regardless of sleep. However, the size congruity effect for the Malay words was only present in the sleep group in Experiment 1 and for word-pairs with strongly encoded (64 presentations) targets and foils in Experiment 3. Applying these findings to CLS models (Davis & Gaskell, 2009; McClelland et al., 1995), McClelland (2013) (reviewed in Chapter 1) highlighted that rapid integration could occur in items that are highly consistent with existing information. Therefore participants from both wake and sleep/nap conditions showed measures of integration for the task-dependent semantic distance effects as they were only required to learn a small set of new words that were directly related to existing high frequency words. In addition, findings from Experiment 3 also suggest that semantic distance automaticity effects can also develop with a limited level of training and encoding. In contrast, for participants who received the same amount of training, the size congruity effect that occurred independently of task demands was present after overnight sleep but not after a period of daytime napping. Therefore, when learning new second language words, with the exception of intense training and encoding (strongly encoded words in Experiment 3), both quantity and quality of sleep seem crucial for size congruity automaticity effects to develop.

Chapter 4: Time-course of Knowledge Integration in New Word Learning

4.1 Introduction

Although there have been multiple studies investigating the benefits of sleep on memory consolidation, there has been less work exploring the time-course of memory consolidation – dissociating consolidation that is time-dependent from consolidation that is sleep-dependent. When considering the time-course of consolidation, it is important to also consider the role of (post-learning) wakefulness. With regard to the term 'consolidation', researchers often adopt the definition mentioned in Chapter 1 where consolidation refers to both an increase in the stability of the memory trace and a qualitative enhancement of the memory trace. Walker (2005) highlighted that considering memory stabilization and memory enhancement as two different outcomes of consolidation may give more insights of the individual role of wakefulness (time) on learning compared to sleep. In the model mentioned by Walker (2005), memory stabilization enables the new memory trace to be preserved and more resistant to forgetting. However, memory stabilization does not lead to any improvement in memory such that it enhances behavioural performance. Rather, performance remains at the same level as initial acquisition. In contrast, memory enhancement produces a qualitative change in the memory trace such that behavioural performance is quantitatively and qualitatively improved from initial learning (Walker, 2005).

According to Stickgold and Walker (2007), memory stabilization of newly learnt knowledge can occur during wakefulness, but memory enhancement relies on changes in brain activity that have been predicted to be dependent on sleep. The above dissociation was depicted in a study by Walker, Brakefield, Morgan, Hobson and Stickgold (2002) where the experimenters trained participants on a procedural finger tapping task. The experimenters found that improvements in participants' RTs after 12 hours of wakefulness were equivalent to the improvements participants would have made if they had undergone additional training. Therefore, the passage of time associated with wakefulness was sufficient to stabilize performance against forgetting and maintaining it at 'training levels'. However, when the 12 hours post-learning included a period of overnight sleep, participants' improvements in the behavioural task was significantly greater than what would have been attributed to additional training. Moreover, additional rehearsal during training and wakefulness (maintenance) benefitted all participants to the same extent whereas for participants who slept, behavioural improvements were further positively correlated to percentage of time spent in NREM Stage 2 sleep. This suggests that even though memory stabilization can occur during wakefulness, qualitative enhancements in memory are sleep-dependent.

Still, results from declarative learning studies have been less conclusive on the effects of wakefulness and sleep on memory consolidation. Payne et al. (2012) trained participants on semantically related and unrelated word-pairs. For one group of participants, testing occurred 12-hours after training, where those who remained awake had a 0.8% performance improvement between test and training whereas those who had a period of overnight sleep had a 6.6% improvement. On first examination, it seemed that a period of wakefulness enabled the stabilization of the memory traces as performance between training and test was maintained, whereas sleep led to an enhancement in performance. However, when looking at the semantically related and unrelated word-pairs separately, participants in the sleep group had improved performance in both sets of word-pairs during testing. In contrast, participants who remained awake had improved performance for semantically related pairs but worse performance for semantically unrelated pairs. Results for the semantically related word-pairs seemed to mirror research by Tse et al. (2007, 2011) (see Chapter 1 for elaboration) where consolidation and integration for newly learnt schemas occurred rapidly. Therefore the time-course for newly learnt items with high semantic relevance to existing knowledge may be more time-dependent and a passage of time in wakefulness may be sufficient for both memory stabilization and enhancement. In contrast, when the newly learnt items are more novel (less related to existing knowledge), sleep is required for the consolidation[14] of the new items.

In general, studies looking at a 12-hour (sleep/wake) or 24-hour training-testing period can be useful in adding insights to the time-course of memory consolidation after initial acquisition. However, research on anterograde amnesia also gives support for the value of examining multiple consolidation nights. As reviewed in Chapter 1 (section 1.5.2), anterograde amnesia refers to memory loss that is specific to recent memories (Scoville & Milner, 1957) and is often caused by hippocampal lesions. These hippocampal lesions not only impair very recent memory, but sometimes also impairs memories that were acquired many weeks

[14] It was not possible to dissociate between effects of stabilization and enhancement for the study.

prior to the lesions (McGaugh, 2000). This implies that memory consolidation and integration of new memories may sometimes involve an extended time-course. Based on the findings on hippocampal lesions, longer-term consolidation studies are needed to enable the examination of the changes in the new memory trace beyond the first consolidation night. Sleep studies involving multiple consolidation nights may also help address the question as to whether sleep has short-term benefits towards memory consolidation (i.e. only after initial acquisition) or if the association between sleep and the time-course of consolidation has long-lasting direct effects (Walker & Stickgold, 2010; Winocur, Moscovitch, & Bontempi, 2010).

One of the central findings from Experiments 1 to 3 was that both the quantity and quality of sleep play essential roles toward consolidating and integrating newly learnt words with existing knowledge. Even though previous studies have demonstrated sleep-associated benefits in lexical competition measures of integration after a single night of consolidation (Dumay & Gaskell, 2007; Tamminen et al., 2010), it may be possible that further offline consolidation would lead to greater enhancements in knowledge integration. In a recent study, Tamminen and Gaskell (2013) used a semantic priming paradigm as a measure of the extent that meanings of newly learnt words have been integrated into the lexicon. Unlike the lexical competition paradigm, it was possible to use the priming paradigm to investigate automaticity of processing semantic information outside of participants' conscious awareness (see masked priming experiment in Tamminen & Gaskell, 2013). Participants were trained on the meanings of 34 novel words and were tested on the same day (Day 1), after a night of consolidation (Day 2) and after a week (Day 8). During the test session, the experimenters used the novel words as primes for a lexical decision task of semantically related existing words. If the novel words were successful primes for the existing words, it could be interpreted that the meanings of the novel words were integrated into the lexicon. Results highlighted that the novel word primes lead to significantly swifter processing of existing words. However, when looking at individual test sessions, the main effect of priming was only significant during the Day 8 test session but not for the other sessions (Day 1 and Day 2). Tamminen and Gaskell's (2013) findings provide important insights into the time-course of semantic integration, suggesting that even though newly learnt words can be integrated into the lexicon rapidly and potentially without sleep, the measure of integration was still stronger after a long period of sleep-associated consolidation.

Another recent new word learning study conducted in children also suggests that declarative learning requires a prolonged time-course that involves multiple nights of offline consolidation (Henderson, Weighall, & Gaskell, 2013). In the above study, children learnt 28 new scientific words and were tested on their memory for the new words on three occasions: immediately after training, 24 hours after training and 1-week after training. Lexical pause detection tasks revealed no significant measures of integration for the new scientific words when the children were tested immediately after training. In contrast, the children displayed sleep-associated knowledge integration whereby lexical competition effects for the new words were present at the 24 hours session. More crucially, when examining explicit recall of the new words, the children were able to recall significantly more words at the 1-week test compared to the 24-hour test. Therefore, even though a single night of consolidation is sufficient to promote measures of knowledge integration, additional (extended) offline consolidation opportunities enabled further enhancements in memory consolidation of the newly learnt words.

Still, an alternate interpretation of Tamminen and Gaskell’s (2013) and Henderson et al.’s (2013) findings could be that a long passage of time, rather than sleep, is sufficient for stronger integration of declarative information. Although Tamminen and Gaskell’s (2013) and Henderson et al.’s (2013) findings do not directly reject the alternative claim, current research that investigates changes in brain activity after learning would suggest that a sleep-associated benefit is a more parsimonious explanation for the gains in the above studies. In Takashima et al.'s (2006) study (mentioned in Chapter 3) the time course of declarative memory consolidation was investigated whereby participants were tested at four time points after learning: days 1 (same day as learning), 2, 30 and 90. There was a positive relationship between SWS experienced during the day 1 nap and declarative memory that was present even 30 days after initial learning. fMRI results also revealed an overall effect of the passage of time from day 1 to 90 where with a longer consolidation period, there was a decrease in participants’ hippocampal activity and an increase in neocortical activity. Applying findings to the CLS model (McClelland, McNaughton, & O’Reilly, 1995), fMRI data supports the notion that gains in declarative memory consolidation continue beyond one night of offline consolidation. By incorporating sleep polysomnography and fMRI in addition to behavioural testing, the experimenters were also able to dissociate between the direct effects of sleep as compared to the overall effects of time on memory consolidation. Spatial overlap between effects of time and the correlations with

SWS highlighted that the duration of SWS was related to changes in hippocampal activity, suggesting that both quantity and quality of sleep play an essential role towards declarative memory consolidation.

4.2 Experiment 4

The first key finding from Experiment 1 was that even though semantic distance effects for newly learnt words were present after a 12-hour training-test interval regardless of sleep, only participants who had overnight sleep displayed a significant size congruity effect for the new words. The next main finding was that a 12-hour duration consisting of overnight sleep enabled greater size congruity and semantic distance effects for newly learnt words as compared to an equivalent time of wakefulness. In addition, the wake group had greater CV_{RTs} for Malay compared to English words. In contrast, there was no significant difference in CV_{RTs} between Malay and English words for the sleep group, suggesting that sleep enhanced automaticity in processing newly learnt words such that they matched the processing of well-established words. Still, it was unclear from Experiment 1 whether further sleep-associated gains would occur if participants were allowed to have additional nights of consolidation.

Experiment 4 aimed to further investigate the effects of time and sleep on the integration of new word meanings with existing knowledge across multiple consolidation nights. As mentioned in the discussion of Experiment 1, sleep seems to be more crucial for measures of automaticity that were related to differences in SRC (i.e. size congruity effect) as compared to measures that were more independent of SRC (i.e. semantic distance effect). Findings from Experiment 1 supported results from previous work (Tham et al., 2011) where participants who had to select the semantically larger word from newly-learnt word-pairs displayed significant size congruity effects in the sleep but not the wake group. In contrast, the semantic distance effect emerged for both sleep and wake groups. Experiment 1 predicted that the semantic distance automaticity effect draws mainly on the strength of specific individual mappings acquired during training whereas the size congruity automaticity effect draws off general and global information integrated across existing knowledge. In summary, it seems plausible that the higher threshold for semantic access needed for newly learnt declarative information, the more crucial the role of sleep as the information would need stronger levels of integration with existing knowledge.

In addition to the semantic comparisons tasks where participants were asked to select the semantically larger word from word-pairs, both Rubinsten and Henik (2002) and Tham et al. (2011) also included physical comparison tasks in their experiments. For the physical comparison tasks, participants were instructed to selected the word with the larger font size and ignore the semantics of the task. Both Rubinsten and Henik (2002) and Tham et al. (2011) found that for well-established words, the size congruity effect was present in the physical comparison tasks.[15] In contrast, both the sleep and wake groups in Tham et al.'s (2011) study did not display a significant size congruity effect for the newly learnt words in the physical comparison tasks. The different results for the newly learnt words in the semantic and physical comparison tasks could be due to the size congruity effect in the physical comparison task being more autonomous and requiring more automatic semantic access than that for the effect in the semantic comparison task. Based on the prediction made in the above paragraph, it thus seems plausible that with multiple nights of consolidation, the size congruity effect would also emerge in the physical comparisons of new words.

In is important to note that the concept of automaticity is a controversial one (Perfetti, 2007). Nonetheless, for some the division between autonomous, intentional and incidental automaticity has been useful (Perlman & Tzelgov, 2006; Rubinsten & Henik, 2002; Tzelgov, Meyer, & Henik, 1992). Autonomous automaticity involves the influence of variables that are not explicitly task-dependent, such as the classic Stroop paradigm (Stroop, 1935). In contrast, intentional automatic processing relates to variables that are part of a global task requirement, such as sentence reading (proficient readers), whereby each word is processed when one reads a sentence to understand what it means (Tzelgov, 1999). Therefore in the semantic comparisons tasks, the semantic distance effect and the influence on the semantic dimension in the size congruity effect are both reflections of intentional automaticity. In contrast, autonomous automaticity is displayed when

[15] Note: Rubinsten and Henik (2002) explained the lack of semantic distance effects in the physical comparisons by arguing that even though the relationships between form-meaning mappings of animal names were mentally represented in a continuum according to magnitude, they were done so in an abstract 'gist-like' manner. Hence, the disparity between meanings in word-pairs did not significantly contribute to differences in reaction times in the physical comparison tasks where they were not directly relevant to performance.

referent-size (semantics) affect physical comparisons (i.e. selecting the larger item based on font size) of the novel word-pairs. Incidental automaticity is defined by swifter processing of the dimensions that were not explicitly taught during acquisition of the new knowledge (Perlman & Tzelgov, 2006). For example, if participants learnt a sequence of words during training and had to read the words as quickly as possible during testing, incidental automaticity occurs when participants are faster if the words are presented in the same order as training than in a different order (Perlman & Tzelgov, 2006). Based on Tzelgov's (1999) review, one of the theoretical differences between both automaticity effects is the level of automatic semantic access required for the effect to be displayed, whereby autonomous automaticity can be interpreted to be a purer measure of automatic semantic access than incidental[16] and intentional automaticity.

Therefore, Experiment 4 aimed to incorporate both physical and semantic comparisons to investigate the time course of sleep-associated memory integration for newly learnt words. Participants were trained on new Malay words using a second language learning paradigm replicated from Experiments 1, 2 and 3. Participants were then tested on how automatically they processed the meanings on the new words by making physical and semantic comparisons of newly learnt word-pairs. In order to explore effects of longer term consolidation and the time-course for integration, participants were tested on 4 sessions starting from the day of training (Day 1, pre-consolidation), followed by 3 consecutive days (Days 2 to 4) separated by overnight consolidation. In Experiments 2 and 3, participants displayed semantic distance effects for newly learnt words not only after a 90-minute nap but also after an equivalent period of wakefulness. Based on findings from Experiments 1 to 3 and previous research by Tse et al. (2007, 2011), it is possible that semantic distance effects may emerge after a shorter training-test duration. Therefore, the first hypothesis was that the distance effect for the newly learnt words would be significant regardless of overnight consolidation (e.g. emerge from Day 1). However, it is unclear whether additional consolidation nights would lead to stronger distance effects. The sleep group in Experiment 1 demonstrated stronger distance effect than the wake group for the newly learnt words (congruent trials). There was also a positive relationship between SWS and the strength of the distance effect. Since the additional consecutive consolidation nights would give participants the opportunity to experience more SWS, it was predicted that even though the

[16] The current designs of the thesis experiments do not allow for the investigation of sleep-associated changes in incidental learning.

semantic distance effect would emerge before consolidation, subsequent sleep-associated consolidation would lead to stronger distance effects for the new words. In addition, the gain in distance effects was hypothesised to be positively related to the quantity of SWS. The above predictions were only made for the semantic comparisons of the new words. It was predicted that for physical comparisons of newly learnt and established words, participants would not display any semantic distance effects. This is because previous work from both Rubinsten and Henik (2002) and Tham et al. (2011) indicated that the distance effects were not present in physical comparisons on well-established words.

In Experiment 1, participants who experienced overnight sleep displayed a significant size congruity effect, whereas those who remained awake for a comparable duration did not display a main effect of congruity. Moreover, participants who had greater sleep spindle activity also displayed stronger size congruity effects (trials with large font differences and semantic distances). It seems that both quantity and quality of sleep were crucial for establishing congruity effects, as participants in Experiments 2 and 3 who took a nap after learning did not display size congruity effects for the new words. Therefore, it was hypothesised that for the semantic comparisons in Experiment 4, participants would only display a significant size congruity effect for the newly learnt words after sleep (i.e. from Day 2) and not during the Day 1 test session. It was also hypothesised that the strength of the congruity effect would increase with each consolidation night[17]. Since the size congruity effect in the physical comparison task was thought to require more automatic semantic access than the effect in the semantic comparison task, and did not emerge after a single night of consolidation (Experiment 1), it was predicted that for the physical comparison tasks, the size congruity effect for new words would require a longer time-course than that in the semantic comparisons, and would probably only emerge after at least two consolidation nights.

As mentioned in Chapter 2, the CV_{RT} can be viewed as a marker of the level of automaticity in processing where a reduction in CV_{RTs} would suggest an increase in automaticity (Segalowitz & Segalowitz, 1993). In Experiment 1, participants who experienced overnight sleep after learning also showed significantly smaller CV_{RTs} for the newly learnt words as compared to participants who remained awake. This suggested that sleep led to increased automaticity in processing new words. It was

[17] As a mobile recording device was used to record sleep activity, it was not possible to extract any sleep spindle information, hence no predictions were made regarding the size congruity effects and sleep spindle activity.

hypothesised that for the semantic comparisons of new words, participants in Experiment 4 would also display smaller CV_{RTs} with each test session. In contrast, it was predicted that there would be no equivalent differences for the established English words. However, for the physical comparisons, it was predicted that there will be no difference in CV_{RTs} with each test session for both the Malay and English words as participants are making explicit judgments about physical font size rather than semantics.

4.2.1 Methods

Apparatus and Materials

The test apparatus consisted of a computer with a 15-inch screen. DMDX software (Forster & Forster, 2003) was used to collect all RT and error data and participants used a USB-joypad for all experimental responses. The sleep polysomnography data was recorded using a wireless sleep monitoring system (ZEO, Axon Laboratories, Newton, MA, USA). The system was able to record sleep and score wakefulness, light sleep, deep sleep and REM sleep in 30 second epochs (Shambroom, Fábregas, & Johnstone, 2012). The system has been validated by external sources (Griessenberger, Heib, Kunz, Hoedlmoser, & Schabus, 2013) and used in experimental studies (Gumenyuk et al., 2011).

Participants were exposed to a total of 18 new Malay words (and their corresponding English animal names) and a separate set of 18 English animal names that were used as baseline word-pairs. All Malay and English words were identical to those from Experiment 2, except that the words were learnt in a single condition instead of different wake and nap conditions. Two separate sets of Malay-English (translation) stimuli were created and participants were randomly allocated to either set of stimuli.

Stimuli & Design

For both training and test sessions, manipulation within each trial was also replicated from Experiment 2. However, participants were exposed to 18 new Malay words, twice as many items as Experiment 2. Therefore, there were some differences in counterbalancing and repetition of the items in Experiment 4. For the training phase of Experiment 2, participants saw each of the 9 animal names (Malay/English translation) 24 times for the 2AFC feedback trials. In Experiment 4, they saw each of the 18 animal names 12 times: 6 times as the correct item and 6 times as foils. Therefore, participants in Experiment 4 were still exposed to the same number of feedback trials (216 trials in total) as Experiment 2.

For the test session, in order to accommodate the additional physical comparison task, twice the amount of words and to avoid a lengthy experiment that might lead to fatigue, it was not feasible to consider all possible permutations of word-pairs or carry out left/right counterbalancing within each block as was done in previous experiments. Instead, in each (physical or semantic) comparison task, participants were presented with words grouped into 2 blocks for each language and the left/right counterbalancing of items was done across the blocks. In each block, there were 72 trials: 36 word-pairs (18 small and 18 large size distance pairs) x 2 (congruity). Hence when combining both English and Malay trials (2 blocks each), participants saw 288 physical comparison trials and 288 semantic comparison trials in total, 'matching' the 576 trials participants saw in Experiment 2. In addition to the 2AFC RT alertness task, other stimuli were adapted from Experiment 3, including the paper-based matching task (adapted to fit the larger number of new words).

A mixed-subjects design was used. The dependent variable was RT for correct responses. There were five within-subjects independent variables namely: day of testing (day1/2/3/4), language of presentation (Malay/English), congruity (congruent/incongruent), semantic distance (small/large), and block of presentation (first/second). The between subject variable was task order (physical comparison first or semantic comparison first). Similar to Tham et al. (2011), performance in the physical and semantic comparisons were analysed separately. Separate analyses were also conducted where proportion of errors was a dependent variable and presented in Appendix 1.4.

Participants

Participants were 23 monolingual native English speakers (6 males, 17 females; *M* = 21, range: 18-32). Participants were then randomly allocated to either the physical comparison first (n = 12) or semantic comparison first group (n = 11). All remaining recruitment procedure was replicated from previous experiments.

Procedure

On the day before the first training session, participants arrived at the testing laboratory to fill up the relevant consent forms. Participants were also shown how to operate the ZEO wireless sleep recording device. They then used the wireless sleep device for a baseline sleep recording during the same night. Prior to the first training session (day 1), participants who did not have a complete sleep recording for the baseline night were re-trained on steps to operate the sleep device. A summary of procedure for the experimental sessions can be found in Figure 13. All participants arrived at the laboratory during approximately the same time for all experimental days. For day 1, participants completed a training session, followed by a paper-based matching task. The final task on day 1 was the test session. For days 2 and 3, participants completed the test session first, followed by the matching task. They were then re-trained on the Malay words. This was done so that performance during the test session would be a more accurate measure for individual differences in sleep-associated memory consolidation that was not contaminated by any practice effects from training. Procedure on the final experimental day (day 4) was similar to days 2 and 3, except that participants completed the WASI IQ test instead of the training session. In addition to the behavioural tasks, participants also used the wireless device to track their overnight sleep patterns for days 1 to 3.

	Day 1	**Day 2**	**Day 3**	**Day 4**
Task 1	Training - Exposure - 2-AFC	Test - Physical - Semantic	Test - Physical - Semantic	Test - Physical - Semantic
Task 2	Matching Task	Matching Task	Matching Task	Matching Task
15 minute Break				
Task 3	Test - Physical - Semantic	Training - Exposure - 2-AFC	Training - Exposure - 2-AFC	WASI IQ

Figure 13. Experiment 4: Overview of tasks for each experimental session. Training sessions are depicted in grey columns and test sessions are depicted in black columns.

Training (grey columns). Before starting the session, the SSS and the 2AFC RT alertness task were administered. For the first part of the training session, participants were exposed to the 18 new Malay words and their relevant English meanings in 18 unique exposure trials. Similar to the previous experiments, the 18 trials were repeated in 3 separate blocks and presented in random orders within each block. For the second part of the training session, participants completed 2AFC feedback trials with the same procedure as Experiment 2.

Test (black columns). The SSS and the 2AFC RT alertness task were first administered. Next, participants completed both physical and semantic comparison tasks where they saw word-pairs consisting of the new Malay words or baseline English words. For the physical comparisons, participants had to ignore the meanings of the words and select the word with the larger font size. For the semantic comparisons, participants were instructed to select the semantically 'larger' item from word-pairs, regardless of the font size. Within each comparison, participants always completed the trials with English word-pairs, followed by the trials with the newly learnt Malay words. The order of the comparison tasks was also counterbalanced such that participants in the physical-first group always completed the physical comparisons first followed by the semantic comparisons, whereas participants in the semantic-first group completed the semantic comparisons before the physical comparisons. The remaining test session manipulation (such as timing of individual trials) was replicated from previous experiments.

4.2.2 Results

There was a significant difference between SSS ratings both training and testing sessions for day 1 [$F(1, 22) = 6.82$, $p = .016$] and a marginally significant difference for day 3 [$F(1, 22) = 3.19$, $p = .022$], where in both instances, participants had higher sleepiness ratings at the session after the 15 minutes break as compared to the session at the start of the experiment day. There were no other significant differences in SSS ratings and 2AFC alertness task performance during training and testing sessions for each experimental day (all p's > .17). There was a main effect of test day for participants' performance on the paper-based matching task [$F(3, 66) =$

5.35, p = .002] where in general, participants performed better with time. Individual comparisons revealed that participants performed significantly better on day 4 as compared to day 1 [F(1, 22) = 5.37, p = .030] and day 2 [F(1, 22) = 7.08, p = .014]. Participants also performed significantly and marginally significantly better on day 3 as compared to day 2 [F(1, 22) = 5.63, p = .027] and day 1 [F(1, 22) = 4.02, p = .058] respectively. There were also no significant differences between participants' performances for the matching task for all other pairs of test days (all p's > .15). All participants also had WASI IQ scores within 3 standard deviations of the general participant and population mean.

Training Session

The mean proportion of errors made for feedback trials across all days was 6.4%. Participants' error rates were arcsine square-root transformed and an ANOVA was performed on error rates with day of training as a within-participant variable. There was no main effect of training day [F(2, 44) = 2.18, p = .13], suggesting that participants were able to learn the individual form-meaning mappings for new words quickly and that their memory for the mappings did not change significantly during the course of the experiment.

Test Session: Semantic Comparisons

Table 15. Experiment 4: Mean response time (in milliseconds) for all correct responses in the semantic comparisons after data trimming. Standard errors are presented in parentheses.

	English				Malay			
	Day 1	Day 2	Day3	Day 4	Day 1	Day 2	Day3	Day 4
Block 1								
Congruity								
Congruent	833	723	665	658	1472	1010	957	832
	(34)	(27)	(21)	(24)	(89)	(52)	(67)	(50)
Incongruent	842	730	685	661	1466	1097	1005	891
	(31)	(26)	(23)	(20)	(84)	(60)	(67)	(49)
Semantic Distance								
Small	877	758	704	693	1563	1121	1040	912
	(36)	(30)	(24)	(23)	(91)	(60)	(71)	(55)
Large	794	694	646	626	1375	986	922	811
	(28)	(22)	(20)	(21)	(82)	(52)	(64)	(44)
Block 2								
Congruity								
Congruent	828	763	705	705	1216	973	915	846
	(34)	(33)	(35)	(34)	(75)	(57)	(73)	(49)
Incongruent	852	773	715	704	1274	977	916	848
	(34)	(31)	(34)	(36)	(75)	(55)	(69)	(48)
Semantic Distance								
Small	886	810	740	754	1314	1022	969	908
	(35)	(34)	(34)	(39)	(77)	(58)	(77)	(53)
Large	794	725	680	655	1175	928	862	786
	(32)	(30)	(36)	(31)	(72)	(51)	(66)	(43)

RTs underwent two levels of trimming to detect and discard outliers. The first trimming procedure was identical to Experiment 2 where RTs for the English and Malay trials were separately trimmed based on presentation block (first vs. second). For the second level of trimming, RTs were also trimmed separately for each day. Trials more than 2 standard deviations away from each participant's mean RT for the corresponding task were discarded and the remaining RTs are presented in Table 15. This resulted in an average data loss of 4.8% and 4.8% per participant for the Malay and English comparisons respectively.

A six-way mixed ANOVA was conducted with day of testing, language, congruity, semantic distance and presentation block as within-participant variables and task order as a between-participant variable. Similar to previous experiments, in the cases where main effects and interactions (where language was a factor) were significant for the combined English and Malay ANOVA, further analysis was conducted to investigate if the main effect and interactions were present in both languages or only in one language. Interactions with presentation block were not presented except for cases of theoretical interest. There was a large main effect of language [$F(1, 21) = 59.17$, $p < .001$], where RTs for the English trials were significantly quicker than those for the Malay trials. A significant main effect of day [$F(3, 63) = 57.27$, $p < .001$ for the combined ANOVA] also emerged for both the English [$F(3, 57) = 26.51$, $p < .001$] and Malay [$F(3, 57) = 57.66$, $p < .001$] trials. According to Figure 14, for both English and Malay trials, participants' overall RTs seemed to be swifter with each test session. In addition, there was a significant interaction between language and day [$F(3, 57) = 26.51$, $p < .001$], which according to Figure 14 suggests that the benefit of multiple nights of consolidation (test day) on improving RTs was greater for the Malay compared to English trials.

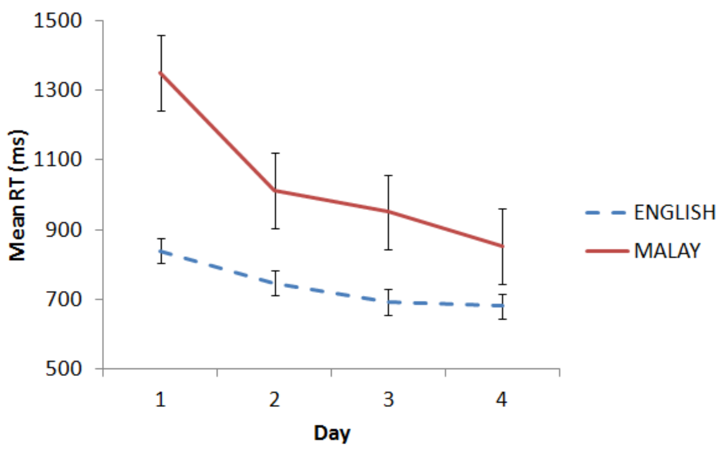


Figure 14. Experiment 4: Mean RTs for English and Malay semantic comparisons across test days. Error bars represent standard error of the means.

The main effect of congruity was significant in the combined ANOVA [$F(1, 21) = 10.81$, $p = .004$] and also when considering both the English [$F(1, 19) = 5.65$, $p = .028$] and Malay [$F(1, 19) = 13.18$, $p = .002$] trials separately (see Figure 15). Even though the size congruity effect seemed visually stronger in the Malay compared to English trials, there was no significant language by congruity interaction [$F(1, 21) = 2.79$, $p = .11$].

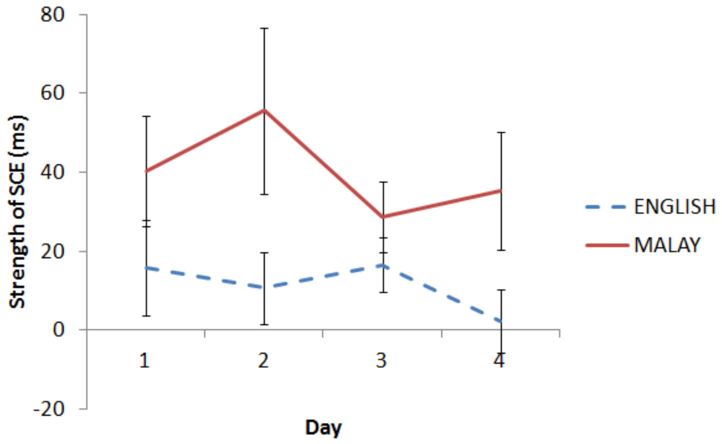
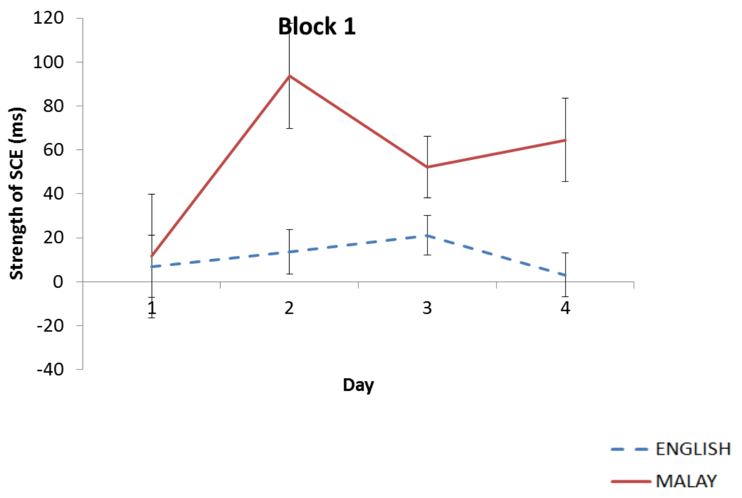


Figure 15. Experiment 4: Strength of Size Congruity Effects (SCE) for English and Malay semantic comparisons across test days. Error bars represent standard error of the means.

There was also an interaction between language, day, block and congruity [$F(3, 57) = 3.00$, $p = .037$]. The higher order interaction was driven by the interaction between day, block and congruity for the new Malay words [$F(3, 57) = 3.30$, $p = .026$] as compared to the English word-pairs [$F(3, 57) = .54$, $p = .66$]. Further analysis revealed that three-way interaction was driven by a two-way interaction between day and congruity in the block 1 Malay trials [$F(3, 57) = 3.21$, $p = .030$]. There was no equivalent interaction for the block 2 Malay trials [$F(3, 57) = 1.38$, $p = .26$], suggesting that the relationship between the day of testing and the strength of the size congruity effect for new words was greater in the first test block (see Figures 16a and 16b). Figure 16a highlights that when looking at block 1 Malay trials on individual days, participants only displayed a significant size congruity effect in day 2 [$F(1, 19) = 15.03$, $p = .001$], day 3 [$F(1, 19) = 14.70$, $p = .001$] and day 4 [$F(1, 19) = 19.63$, $p < .001$], which all occurred after overnight consolidation. There was no main effect of congruity for the block 1 Malay trials experienced on day 1 [$F(1, 19) = .011$, $p = .92$] where participants were tested before overnight consolidation. In order to investigate the effects of time and sleep-related consolidation on the congruity effect, pair-wise comparisons were carried out between the consecutive days, where p-values were Bonferroni corrected for the number of comparisons made. This resulted in a threshold p-value of .0167 after Bonferroni corrections. Pair-wise comparisons revealed that participants had significantly stronger size congruity effects in the day 2 compared to day one block 1 test sessions for the Malay word-pairs [$F(1, 21) = 7.36$, $p = .013$]. In contrast, even though participants also displayed a trend towards stronger size congruity effects between day 2 compared to day 3 [$F(1, 21) = 7.36$, $p = .092$] and also on day 4 compared to day 3 [$F(1, 21) = 7.36$, $p = .51$], both the effects were not statistically significant.

(a)



(b)

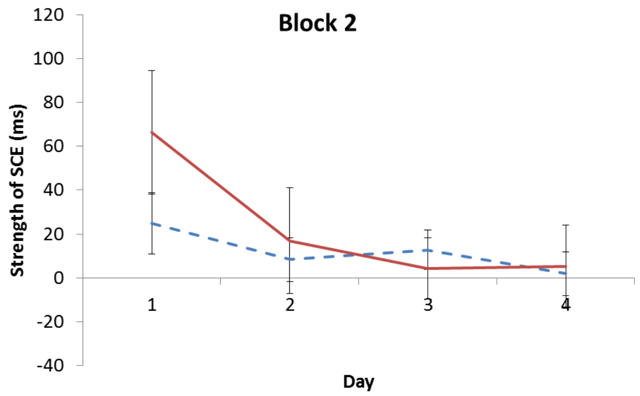


Figure 16. Experiment 4: Strength of Size Congruity Effects (SCE) for English and Malay semantic comparisons across test days in presentation (a) Block 1 and (b) Block 2. Error bars represent standard error of the means.

There was a significant main effect of semantic distance in the combined ANOVA [$F(1, 21) = 166.35$, $p < .001$] and also when separating both the English [$F(1, 19) = 184.55$, $p < .001$] and the Malay [$F(1, 19) = 135.36$, $p < .001$] trials (see Figure 17). There was an interaction between language and distance [$F(1, 21) = 24.66$, $p < .001$], whereby the main effect of distance was stronger in the Malay as compared to English trials. Similar to the size congruity effect, the stronger semantic distance effect in the Malay compared to English trials were potentially due to the fact that the Malay words were matched for word length but the English word-pairs were not matched for word length. The difference in word-length between the English word-pairs may have affected participants' processing of the semantic differences between the English words (Dehaene, 1996; Moyer, 1973; Rubinsten & Henik, 2002). There was also an interaction between day and distance ANOVA [$F(3, 63) = 4.11$, $p = .010$]. Similar to the size congruity effects, pair-wise comparisons were also conducted on consecutive days, where p-values were Bonferroni corrected for the number of comparisons made. However, unlike the congruity effects, there were no significant pair-wise comparisons between consecutive days of testing for the semantic distance effects (all corrected p's > .96).

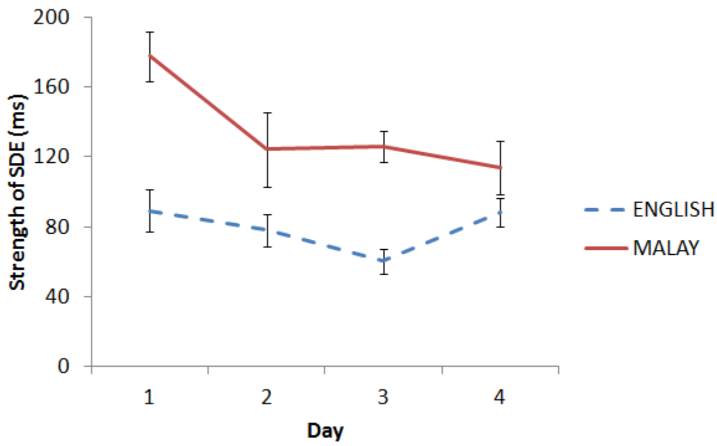


Figure 17. Experiment 4: Strength of Semantic Distance Effects (SDE) for English and Malay semantic comparisons across test days. Error bars represent standard error of the means.

There was a significant effect of block [$F(1, 21) = 6.54$, $p = .018$] and an interaction between language and block [$F(1, 21) = 67.29$, $p < .001$]. Although both the English [$F(1, 21) = 5.12$, $p < .05$] and Malay [$F(1, 21) = 34.93$, $p < .001$] items displayed a main effect of block, participants were quicker to respond in block 1 as compared to block 2 items in the English comparisons whereas the opposite effect of a speed-up of RTs in block 2 compared to block 1 was present in the Malay trials. There was also a three-way interaction between language, day and block [$F(3, 63) = 4.63$, $p = .005$]. When looking at the languages separately, the interaction between day and block was significant in the Malay [$F(3, 57) = 9.82$, $p < .001$] but not the English [$F(3, 57) = 1.64$, $p = .19$] trials. Further analysis of the Malay trials revealed that the speed-up of RTs in block 2 compared to block 1 was significant in day 1 [$F(1, 19) = 24.52$, $p < .001$] and day 2 [$F(1, 19) = 8.69$, $p = .008$]. However, there was no main effect of block for the Malay trials in day 3 [$F(1, 19) = 2.49$, $p = .13$] and day 4 [$F(1, 19) = .63$, $p = .44$] of testing. There were no other significant interactions for the within-participant variables in the semantic comparison RT analysis. There was also no main between-participant task order effect [$F(1, 21) = .46$, $p = .51$].

In order to dissociate between a general speed-up in RT and RT changes due to increased automatic processing, an ANOVA was conducted on participants' CV_{RTs} with language and (testing) day as within-participant variables. In contrast to the RT, analysis, there was no main effect of day [$F(3, 63) = .006$, $p = .99$] for the CV_{RT} analysis. However, similar to the RT analysis, there was a main effect of language [$F(1, 21) = 68.67$, $p < .001$] for the CV_{RT} analysis where the newly learnt Malay items displayed greater CV_{RTs} (less automatic access) than the well-established English items. There was also an interaction between language and day [$F(3, 63) = 6.44$, $p < .001$] (see Figure 18). When comparing CV_{RTs} between languages for the individual test days, participants displayed significantly greater CV_{RTs} (less automatic access) for the Malay compared to English words in day 1 [$F(1, 21) = 88.63$, $p < .001$], day 2 [$F(1, 21) = 25.88$, $p < .001$] and day 3 [$F(1, 21) = 23.35$, $p < .001$]. However for day 4, there was only a marginal significant difference in participants' CV_{RTs} between the English and Malay words [$F(1, 21) = 4.32$, $p = .05$].

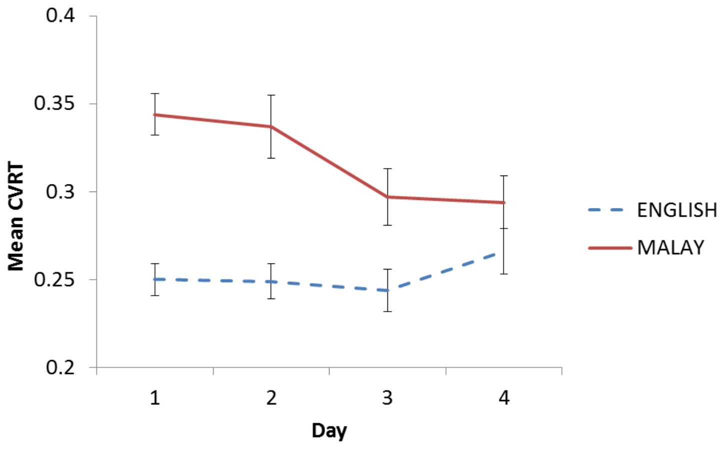


Figure 18. Experiment 4: Mean CV_{RTs} for English and Malay semantic comparisons across test days. Error bars represent standard error of the means.

As there were interactions between language and day for both RT and CV_{RT} analyses, an ANOVA was conducted on participants' RT variance, measured by standard deviations in RT (SDs). There was also a main effect of language [$F(1, 21) = 6.97$, $p = .015$] and a significant interaction between language and day [$F(3, 63) = 11.10$, $p < .001$]. This suggests that the findings for the CV_{RT} analysis was not driven only by practice effects (changes in mean RT), but were also driven by the reduction in RT variance.

Test Session: Physical Comparisons

Table 16. Experiment 4: Mean response time (in milliseconds) for all correct responses in the physical comparisons after data trimming. Standard errors are presented in parentheses.

	English				Malay			
	Day 1	Day 2	Day 3	Day 4	Day 1	Day 2	Day 3	Day 4
Block 1								
Congruity								
Congruent	351	345	331	344	348	344	329	343
	(5)	(12)	(4)	(10)	(5)	(11)	(4)	(12)
Incongruent	369	347	336	343	347	363	331	340
	(8)	(10)	(5)	(8)	(7)	(25)	(4)	(13)
Semantic Distance								
Small	361	348	334	342	347	349	331	346
	(7)	(11)	(4)	(8)	(6)	(14)	(4)	(15)
Large	360	344	333	344	348	358	329	338
	(6)	(11)	(4)	(10)	(6)	(22)	(4)	(11)
Block 2								
Congruity								
Congruent	348	342	339	334	344	346	330	327
	(7)	(7)	(4)	(6)	(6)	(8)	(5)	(5)
Incongruent	349	344	334	336	336	344	328	334
	(6)	(9)	(4)	(7)	(5)	(7)	(5)	(7)
Semantic Distance								
Small	348	343	337	335	341	344	329	329
	(6)	(8)	(4)	(6)	(5)	(7)	(5)	(6)
Large	350	343	336	335	339	346	329	332
	(7)	(8)	(4)	(6)	(6)	(8)	(5)	(7)

RTs underwent two levels of trimming to detect and discard outliers that were identical to the semantic comparisons. This resulted in an average data loss of 5.4% and 5.3% per participant for the Malay and English comparisons respectively for the whole experiment. The remaining RTs were presented in Table 16. A six-way ANOVA was performed with identical factors as the semantic comparisons.

Interactions with presentation block were not reported except for cases of theoretical interest.

There was no significant main effect of language [$F(1, 21) = 2.74$, $p = .11$] and there was a marginally significant effect of day [$F(3, 63) = 2.52$, $p = .066$]. There was a significant interaction between language and day [$F(3, 63) = 3.50$, $p = .021$] (see Figure 19), which was driven by the notion that participants had faster RTs in selecting the word with the larger font size for the Malay word-pairs as compared to the English word-pairs on the first day of testing [$F(1, 21) = 11.81$, $p = .002$]. In contrast, there was a marginally significant difference in RTs between Malay and English trials for day 2 [$F(1, 21) = 4.17$, $p = .054$], and no significant differences for day 3 [$F(1, 21) = 2.71$, $p = .11$] and day 4 [$F(1, 21) = .28$, $p = .60$] of testing.

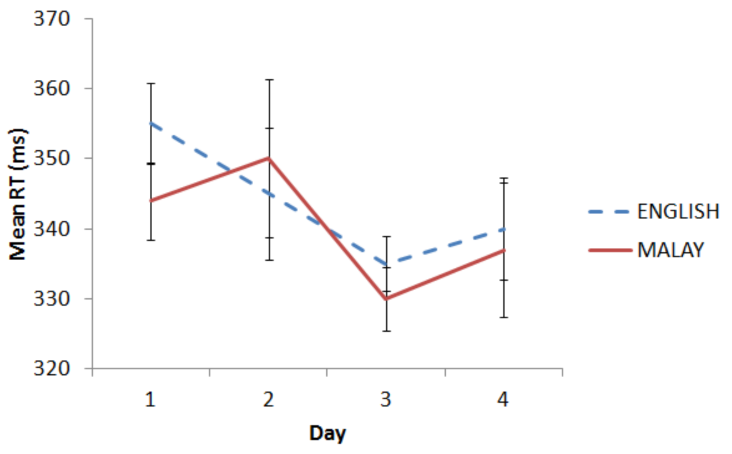


Figure 19. Experiment 4: Mean RTs for English and Malay physical comparisons across test days. Error bars represent standard error of the means.

There was no significant main effect of semantic distance [$F(1, 21) = .019$, $p = .89$] or size congruity [$F(1, 21) = 1.31$, $p = .27$]. However, there was a significant interaction between block and congruity [$F(1, 21) = 7.22$, $p = .014$], where participants displayed a marginally significant main effect of congruity word-pairs in block 1 [$F(1, 21) = 4.18$, $p = .054$], but not block 2 [$F(1, 21) = .017$, $p = .090$]. Even though there was no significant interaction between block, language and congruity, the block 1 English and Malay trials were further analysed separately as previous studies revealed that main effects of congruity were displayed for existing English word-pairs (Rubinsten & Henik, 2002; Tham et al., 2011). Further analysis was conducted to determine if the lack of congruity effects in the English trials was due to confounds in stimuli manipulation or related to block effects where extended repetition reduces the strength of the size congruity effect (see section 2.1 of Foltz, Poltrock, & Potts, 1984). When separate analyses were conducted for block 1 English and Malay trials, results highlighted a main effect of congruity that was significant in the English [$F(1, 21) = 9.38$, $p = .006$] but not the Malay [$F(1, 21) = .80$, $p = .38$] trials.

There was also no significant main effect of presentation block [$F(1, 21) = 1.53$, $p = .23$]. All other interactions with the within-participant variables were not significant except the interaction between day, congruity and distance [$F(3, 63) = 3.79$, $p = .014$]. The above interaction was not expanded as further analysis on individual two-way interactions did not yield any theoretically relevant findings. The between-participant main effect of task order was not significant [$F(1, 21) = .023$, $p = .70$].

Similar to the semantic comparisons, in order to dissociate between a general change in overall RT and RT changes due to increased automaticity in processing, an ANOVA was conducted on participants' CV_{RTs} with language and (testing) day as within-participant variables. There was no main effect of language [$F(1, 21) = .69$, $p = .41$] or day [$F(3, 63) = 1.78$, $p = .16$]. However, there was a significant interaction between language and day [$F(3, 63) = 4.57$, $p = .006$] (see Figure 20), where participants' demonstrated smaller CV_{RTs} for English compared to Malay words on day 2 [$F(1, 21) = 5.23$, $p = .032$]. In contrast, there was no difference in participants' CV_{RTs} between English and Malay words for day 1 [$F(1, 21) = 2.17$, $p = .16$], day 3 [$F(1, 21) = .94$, $p = .34$] and day 4 [$F(1, 21) = .85$, $p > .37$].

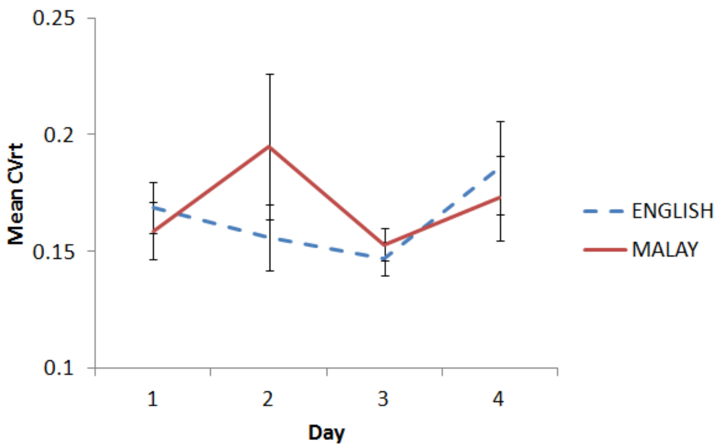


Figure 20. Experiment 4: Mean CV_{RTs} for English and Malay physical comparisons across test days. Error bars represent standard error of the means.

Similar to the physical comparisons, as there were interactions between language and day for both RT and CV_{RT} analyses, an ANOVA was conducted on participants' RT variance for the physical comparisons. There was also a main effect of language [$F(1, 21) = 46.79$, $p < .001$] and a significant interaction between language and day [$F(3, 63) = 53.34$, $p < .001$]. Therefore like the semantic comparisions, the findings for the CV_{RT} analysis in the physical comparisons were likely to be driven by the reduction in participant's RT variance (in addition to general practice effects).

Sleep data from the wireless sleep recording device

Out of the 23 participants, 16 participants had all four nights of sleep recordings, 4 participants had three recorded nights and 3 participants had two or less recorded nights. The main reason for unsuccessful recordings was due to the headband from the wireless device slipping off during the recording night. Out of 92 recorded nights, there were 13 unsuccessful sleep recordings, which amounted to a 14% data loss in total. The remaining sleep parameters are displayed in Appendix 6.

A univariate ANOVA was performed on TST with recording night (of sleep) as a between-participant factor. There was no main effect of recording night [$F(3,75) = 1.16$, $p = .33$]. Post-hoc tests also revealed that there were no significant differences in TST when comparing individual nights (all p's >.40). Next, a repeated-measures ANOVA was conducted with stage of sleep experienced (light, deep, REM) as a percentage of TST[18] as within-participant factors and recording night as a between-participant factor. There was a main effect of sleep stage [$F(2,150) = 202.31$, $p < .001$]. The interaction between sleep stage and recording night was not significant [$F(6,150) = .97$, $p = .45$].

Similarly to Experiment 2 and 3, correlations between sleep parameters and behavioural data did not reveal any significant relationships (all p's > .11), and were hence not expanded upon. It is also important to note that it was not possible to conduct any spindle analysis based on the data obtained by the wireless sleep recording system.

[18] Each sleep stage was analysed as a percentage of TST as there was no main effect of TST for each recording night. The pattern of results is similar when consider the absolute time spent in each sleep stage.

4.3 Chapter Summary and Discussion

Table 17. Experiment 4: Summary of main experimental findings

	Semantic Comparisons	Physical Comparisons
Size Congruity Effect	• Significant main effect for overall ANOVA • Malay word-pairs: For trials in the first presentation block, the size congruity effect was only present from Day 2 (after overnight consolidation)	• No significant main effect across all trials • English word-pairs: The size congruity effect was significant for trials in the first presentation block
Semantic Distance Effect	• Significant main effect for overall ANOVA	• No significant semantic distance effect
Overall RTs	Language x Day: • The effect of test day/consolidation on improvements in RTs were significantly greater for the Malay word-pairs	Language x Day: • Day 1 only: Participants had faster RTs for Malay compared to English trials • Days 2, 3 and 4: No difference in RTs between languages
CV_{RTs}	• Significant effect of day in the Malay but not English trials Test days 1, 2 and 3: • Malay words exhibited significantly greater CV_{RTs} than the English words Test day 4 • No significant difference in CV_{RTs} between Malay and English words	• The main effect of day was not significant across languages Test day 2: • Malay words exhibited significantly greater CV_{RTs} than the English words Test days 1, 3 and 4: • No significant difference in CV_{RTs} between Malay and English words
Sleep Parameters	• No significant correlations between behavioural findings and sleep parameters	

Experiment 4 aimed to investigate the time-course for semantic integration of new word meanings into existing knowledge by examining changes in size congruity effects and semantic distance effects in physical font and semantic (referent) comparisons of newly learnt Malay word-pairs over four days. Summaries of the main experimental findings are depicted in Table 17.

Semantic comparisons: size congruity effects

One of the main findings was that for semantic comparisons of Malay word-pairs occurring in the first presentation block, the size congruity effect was only present in the sessions after at least one night of post-learning consolidation had occurred (i.e. from Day 2). This supported the first experimental hypothesis that for the semantic comparisons, participants would only display a significant size congruity effect for the newly learnt words after sleep (i.e. from Day 2) and not during the Day 1 test session. The current finding provides support for Tham et al.'s (2011) study and Experiment 1, as both experiments also highlighted that size congruity effects in new words were only present after participants experienced overnight sleep but not after an equivalent period of wakefulness. In Experiments 2 and 3 (presented in Chapter 3), it was predicted that the lack of size congruity effects for the newly learnt word-pairs was due to the fact that a daytime nap (maximum 90 minutes) was not sufficient for enhancing the quality of knowledge integration in new word learning. The current pattern of findings supports the above prediction by implying that a minimum duration of sleep is needed for the size congruity integration measure to occur.

It was also hypothesised that the strength of the congruity effect for the Malay words would increase with each consolidation night. The pattern of results presented in Figure 16a suggested that for the semantic comparisons, there was a trend whereby participants displayed stronger size congruity effects with each testing session for block 1 of Malay trials, supporting the hypothesis to a large extent. Interactions between size congruity and test days highlighted that out of all consecutive test days (for Malay block 1 trials), only the significant difference in the strength of the size congruity effect was between day 1 and day 2. Therefore, even though the strength of the size congruity effect increases with each consolidation night, the first consolidation night after learning seems to be the most crucial in benefitting measures of knowledge integration for newly learnt words.

Initial examination of the size congruity effects for the Malay words in Experiment 4 might raise concerns as to why the autonomous automaticity effects were only significant in the first presentation block. It should be noted that such block effects were also found in previous research that investigated the size congruity paradigm, such as the study by Foltz et al (1984) that was mentioned in Chapter 2 (section 2.1). In one condition of their study, Foltz et al (1984) showed participants 192 experimental trials with unique word-pairs (i.e. each word-pair was only viewed once). The experimenters then separated participants' responses into four blocks. Results revealed that the size congruity effect was strongest in the first block and that the strength of the size congruity effect diminished with each presentation block. In a separate condition, Foltz et al (1984) also found that when participants were shown word-pairs that were repeated over multiple trials, the size congruity effects diminished rapidly. Therefore it is not surprising that in the current experiment where participants viewed word-pairs that were repeated, crucial findings regarding the size congruity effect were only present in block 1.

Semantic comparisons: semantic distance effects

It was predicted that for the semantic comparisons, the distance effect for the newly learnt words would be significant regardless of consolidation (e.g. emerge from Day 1). As hypothesised, participants displayed a significant distance effect for the newly learnt Malay words throughout the experiment. The distance effect was also present in the day 1 test session that occurred approximately 15 minutes after training. This finding was consistent with Experiments 1 to 3 where semantic distance effects emerged for the newly learnt words without overnight consolidation and schema research by Tse et al. (2007, 2011) as mentioned in Chapter 1. In a recent study by van Kesteren, Rijpkema, Ruiter and Fernández (2013), participants were trained on association between visual patterns and object (written word)-fabric item pairs. For example, participants saw a visual pattern on the screen that was also presented with the word 'tie' and at the same time were given a fabric sample to touch. The object-fabric pairs were either semantically consistent (e.g. a fabric that is used for making ties) or inconsistent (e.g. a fabric that is not used for making ties). It is expected that for well-learnt items, there would a benefit of semantic consistency in participants' responses. Participants were allocated into three groups that were tested immediately after, 20 hours after or 48 hours after training. One of the test tasks examined associative memory where participants were shown the visual patterns they saw during training and had to select the correct object (out of three choices) that was presented with the pattern during training. At the test which occurred immediately after training, participants displayed significantly better performance for the semantically consistent, compared to the semantically inconsistent, associations during the associative memory task. Moreover, there were no differences in semantic consistency benefits between the test performance tested immediately after training as compared to the performance 20 hours after or 48 hours after training. The above findings by van Kesteren et al (2013) and the current experimental findings suggest that it is possible for learning and knowledge integration to occur rapidly prior to overnight consolidation.

Still, it is also interesting that even though participants were responding to the same word-pairs in the semantic comparisons, the size congruity and semantic distance effects had different relationships with overnight consolidation. It seems that the size congruity and semantic distance effects show different levels of sleep-dependence for knowledge integration in new word learning whereby as predicted in Figure 21, the size congruity effects are more reliant than the semantic distance effects on sleep. As highlighted in the preceding paragraphs, the semantic distance effect for the new Malay words was present after training whereas the size congruity effects (block 1) only emerged after overnight consolidation.

<u>Differences in time-course for the emergence of size congruity and semantic distance effects</u>

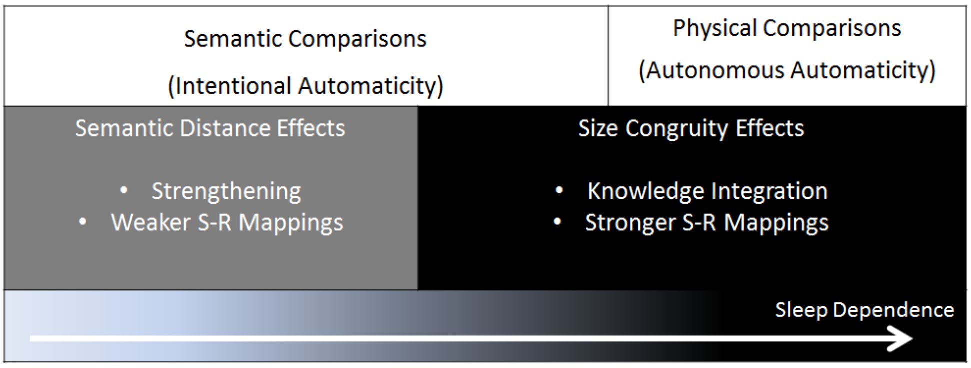


Figure 21. Observed time-course and sleep dependence for semantic distance effects and size congruity effects to emerge in newly learnt words

It is possible that the different relationships with sleep are driven by the difference in 'task complexity' between size congruity effects and semantic distance effects. The question as to whether task complexity affects overnight memory consolidation has been previously addressed by Kuriyama, Stickgold and Walker

(2004). The experimenters trained four groups of participants on a set of motor learning tasks that differed in complexity. The objectives of the task also varied in terms of difficulty for each participant group. Despite the variation in task complexity, there were no between-group differences in the effects of practice on performance. In contrast, participants in the group with the most difficult task objective displayed the largest training-test improvements after overnight consolidation. Moreover, within each of the groups, tasks that were most complex showed the greatest sleep-dependent improvements whereas there were no equivalent benefits of sleep towards less complex tasks. The sleep-dependent between and within-group differences in the study suggest that the benefit of sleep on memory consolidation would be larger for more complex and difficult tasks. Since the size congruity automaticity effect is affected by requires stronger S-R mappings, it can be seen as more complex and difficult than the semantic distance effect. Therefore, Kuriyama et al.'s (2004) findings would predict that sleep-dependent improvements would be larger for the size congruity effect as compared to the semantic distance effect as reflected in the current experimental results.

Another study that supports task complexity in driving differences in benefits of consolidation on the size congruity and semantic distance effects is the work by van Kesteren et al.'s (2013) that was mentioned in the previous paragraph. Apart from the test on associative memory, the experimenters also tested participants' memory for individual visual patterns. Participants were shown visual patterns and had to judge the patterns on two levels, first being whether the patterns were previously taught during training and second, they had to rate how sure (not sure, nearly sure or sure) they were of their choice. In the associative task, participants could use the visual patterns as context cues for the object words, whereas there were no context cues present for the test task. Therefore the item task is considered more complex than the associative memory task. Unlike the associative memory task, the item memory task did not show any benefits of semantic consistency at the immediate task point but benefits of semantic consistency were displayed at the 20 hours and 48 hours test point where participants had experienced overnight sleep. In both the current study and van Kesteren et al.'s (2013) study, only the task that was more complex or demanding benefitted from overnight consolidation.

The differences in the time course of consolidation and task complexity in both the current experiment and van Kesteren et al.'s (2013) study is consistent with the predictions from the general systems consolidation accounts (Diekelmann &

Born, 2010a; Frankland & Bontempi, 2005; McClelland, McNaughton, & O'Reilly, 1995) and the schema learning accounts(Tse et al., 2007, 2011; van Kesteren et al., 2012). Both the semantic distance effect (current experiment) and the associative memory task in van Kesteren et al.'s (2013) study require participants to integrate newly learnt representations into their existing neocortical knowledge. The system consolidation accounts predict that in general, sleep-associated consolidation would be essential for the above integration process. However, it is also important to highlight that representations from both studies are highly related to participant's existing knowledge. In van Kesteren et al.'s (2013) study, the manipulation of object-fabric item pairs was directly related to common everyday items. Although participants were learning a new language in the current experiment, the new Malay words were directly related to high frequency English words and the Malay-English (semantic) pairings were reinforced during training. Therefore, the above findings are not contradictory to systems consolidation accounts. This is because as mentioned in Chapter 1, the adapted CLS model McClelland (2013) predicts that neocortical learning can occur rapidly and without sleep if the new knowledge related to the tasks or paradigms were highly relevant with existing knowledge.

In contrast to the semantic distance effect (current study) or the associative task (van Kesteren et al.'s study), both the size congruity effect (current study) and the item recognition task (van Kesteren et al.'s study) may need to rely on neurocognitive processes beyond the integration of new representations with existing knowledge. As the level of automatic semantic access required to elicit the size congruity effect (in the semantic comparisons) is greater than that for the semantic distance effect, integration of the new Malay words into existing knowledge may not be sufficient to elicit the effect. In order for participants to be affected by the disparity between the font and referent size of the Malay items, they would need to have formed generalized associations between properties (regarding animal sizes) of the new Malay representations that are independent from the English representations (Walker & Stickgold, 2010). Similarly, for the item recognition task in van Kesteren et al.'s (2013) study, participants needed to abstract their memory for the visual patterns without the object-fabric cues that are relevant to existing knowledge. Therefore, since the size congruity effect and item memory task are related and consistent with existing knowledge, current findings are aligned with the systems consolidation account predicting that sleep-associated gains would occur from the overnight transfer of information from the hippocampal to neocortical system (Frankland & Bontempi, 2005; McClelland, McNaughton, & O'Reilly, 1995; McClelland, 2013).

Effects of sleep parameters

Based on findings from Tham et al. (2011) and Experiment 1, it was also hypothesised that the additional SWS from multiple consolidation nights in Experiment 4 would lead to stronger distance effects for the new Malay words, i.e. participants would experience stronger distance effect for each test day. In addition, as there was a positive relationship between SWS and semantic distance effects in Experiment 1. It was also hypothesised that the gains in distance effects for each test day would to be positively related to the quantity of SWS experienced on the previous consolidation night. In contrast to the above hypotheses, there was no strengthening of the semantic distance effect with each test day. Moreover, there was also no relationship between sleep parameters and behavioural effects. However, it is very unlikely that the lack of correlations with sleep parameters was due to confounds related to sleep quality or quantity of specific recordings as ANOVAs revealed that there was no difference between portion of time spent in each sleep stage throughout the experiment. One reason for this difference is that even though both Experiment 1 and Experiment 4 examined relationships with overnight consolidation, there was an approximately 10-hour interval between the training and test session in Experiment 1 (and Tham et al (2011)). In contrast, for Experiment 4, the first test session occurred 15 minutes after training, followed by 3 subsequent test sessions each separated by 24 hours. Therefore prior to the first consolidation night, participants in Experiment 4 would have already been exposed to one test session or seen 288 more exposures of Malay word-pairs than participants in Tham et al (2011) and Experiment 1. After each test session, participants were also re-trained on all the Malay words. It is likely that with the additional exposure and learning opportunities, participants in Experiment 4 would have more robust encoding of the new Malay words. However, recent studies have highlighted that sleep may preferentially benefit weakly encoded items over strongly encoded items. As mentioned in Experiment 2, the nap associated benefits towards the strength of semantic distance and size congruity effects for the new Malay words were significantly greater for participants who had poorer performance during training (less well-encoding new words) than for participants who displayed good encoding of the new words during training. In addition, Drosopoulos, Schulze, Fischer and Born (2007) found that performance for weakly encoded items in a paired associate task was better after overnight sleep compared to an equivalent period of wakefulness, whereas there was no equivalent difference in benefit of sleep for strongly encoded items (see Chapter 1 and Chapter 3 for a more detailed description of the study). As mentioned in Chapter 3, Stickgold (2009) suggested that sleep would benefit moderately encoded items to the greatest extent as very strongly encoded items may not need further sleep-dependent consolidation and very weakly encoded items may be forgotten prior to sleep, whereas moderately encoded items leave room for sleep-dependent gains to emerge. It is therefore possible that sleep related enhancements in the semantic distance effect were present for the new words in Experiment 1 and Tham et al (2011) as they were moderately encoded, but not present for new words in Experiment 4 as they were very strongly encoded prior to overnight consolidation.

It should also be noted that correlations between sleep parameters in Experiment 1 only occurred for behavioural effects that exhibited ANOVA interactions with sleep, whereby the strength of the size congruity and semantic distance effects in Malay words were positively related to sleep spindle activity and SWS respectively. In the current experiment, only the size congruity effects in the Malay semantic comparisons displayed interactions with sleep (i.e. test day). However, as it was not possible to conduct spindle analysis on the data from the recordings from the wireless sleep device, no concrete conclusions can be drawn as to whether participants who experienced greater sleep spindle activity would also show greater gains in the size congruity effects over multiple test sessions. As the size congruity effects seem to be a more sensitive measure of sleep-dependent integration of new words, it may be useful to replicate the current study using measures that will enable spindle activity to be extracted.

Despite the convenience and comfort of sleeping in one's own bed instead of a complete laboratory based polysomnography set-up, there are also limitations to the reliability of a wireless sleep recording device. Griessenberger et al. (2013) highlighted that compared to expert scorers, wireless sleep recording devices had poor reliability in detecting wakeful epochs. This specifically affected the detection of sleep onset latencies, wake time after sleep onset and thus implicated the amount of total sleep time scored. In addition to the poor reliability in detecting wakeful epochs, Tonetti et al. (2013) highlighted that wireless sleep recording devices also overestimated the total amount of REM sleep experienced by participants. Therefore the wireless sleep recording device was less reliable compared to polysomnology. Still despite the issue of detecting wakeful epochs and REM sleep, both studies also indicated the agreement between the wireless sleep recording devices and polysomnography was moderate to high, especially for NREM sleep which is most relevant to the current experiment. Hence, it is unlikely that the lack of relationships with sleep parameters were confounded by the use of the wireless devices to record sleep.

Physical comparisons

In addition to examining semantic comparisons of newly learnt word-pairs, Experiment 4 also investigated effects of time and sleep on the integration of new words by comparing physical comparisons of Malay word-pairs where participants had to ignore the meanings of the words and select the item that is larger in font size. Rubinsten and Henik (2002) and Tham et al. (2011) found that for well-established words in participants' native language, the size congruity effect was present in the physical comparison tasks. In contrast there was no size congruity for physical comparisons of new words in Tham et al.'s (2011) study. It can be argued that the size congruity effect in physical comparisons of new words would represent a greater level of autonomous automaticity than the semantic comparisons (see section 4.2) and hence may need more than one consolidation night for effects to emerge. It was hypothesised that in Experiment 4, the size congruity effect would also emerge in the physical comparisons of new words after multiple consolidation nights. In contrast to the experimental hypothesis, there was no significant size congruity effect in the Malay trials with three post-learning consolidation nights. As size congruity effects were present in the English word-pairs (block 1), the lack of an effect in the Malay trials is unlikely to be due to any circadian or fatigue confounds.

According to some models of systems consolidation, long term consolidation and integration for certain knowledge can occur over many months and even many years (McClelland et al., 1995). As depicted in Figure 21, since the size congruity effect in the physical comparisons relies on greater autonomous processing of semantic information than the semantic comparisons, it seems parsimonious to predict that there would be a difference in time courses for the development of congruity effects in physical as compared to semantic comparisons. The above prediction was highlighted in numerical cognition work by Rubinsten, Henik, Berger, & Shahar-Shalev (2002). Three groups of participants (6-7 year olds, 7-11 year olds and university students) were tested on physical and semantic comparisons of single digit numeral-pairs that were manipulated for congruity and distance. One of the main experimental findings was that for 6-7 year olds, participants displayed a significant congruity effect for the semantic comparisons but not for the physical font size comparisons. The results were not due to task confounds as there were no significant differences in congruity effects between the physical and semantic comparisons for both the 7-11 year olds and university student groups. Therefore, the results revealed that the size congruity effects in the physical comparisons take longer to emerge than those in the semantic comparisons. Drawing from Rubinsten et al.'s (2002) study, it seems plausible that three consolidation nights were not sufficient for size congruity effects to occur for the newly learnt Malay words in the physical comparisons for Experiment 4.

Even though the level of training was sufficient to elicit automaticity effects for the semantic comparisons, it is also possible that the training was not adequate enough for inducing automaticity effects in the physical comparisons. Rodd et al. (2012) highlighted that the level of semantic encoding was crucial in enabling participants to successfully integrate new meanings in existing words. In Rodd et al.'s (2012) third experiment, participants were trained on new meanings for 72 existing words where half of the new meanings were trained using methods that focus on the surface encoding of the existing word form and the new meaning. The other half of the new meanings were trained using methods where participants had a deeper level of semantic encoding of the new meanings and existing words. The experimenters found that after four days of training, participants had swifter responses towards words that were trained with a deep level of semantic encoding compared to surface encoding. Even with four nights of consolidation, the level of semantic coding still moderated processing of the new meaning-word associations. Therefore, it is plausible that if the level of semantic coding was not adequate, extended offline consolidation opportunities may still not be sufficient to induce automaticity effects in the physical comparisons. Future experiments should also examine the interaction between level of semantic encoding and amount of consolidation for the emergence of size congruity effects for new words in the physical comparisons.

Still, physical comparisons of the Malay word-pairs showed that participants experienced other sleep-associated gains apart from the size congruity and semantic distance effects. When considering participants' overall RTs, participants had significantly faster RTs for the Malay compared to English trials during the first test day. In contrast, there was no significant difference in participants' RTs for the Malay and English trials in the subsequent test sessions. This pattern of findings may seem contradictory as participants' should be more familiar with the English words. Moreover, the pattern of results from the semantic comparisons in previous experiments would predict that participants would be faster in processing the English word-pairs. However, it is important to note that for the semantic comparisons, increased familiarity with word meanings would facilitate the decision in selecting the semantically larger item. In contrast, the physical comparison task requires participants to inhibit their knowledge regarding the semantics of the word-pairs and select the item in larger font. Therefore, the most plausible explanation for the current findings would seem to be that during the pre-consolidation test session in day 1, participants were more able to ignore the meanings of the Malay words compared to the well-integrated English words. However, in subsequent post-consolidation test sessions, the lack of difference between English and Malay words suggests the meanings of the new Malay words were better consolidated. In addition, neither the English or Malay word-pairs displayed significant changes in accuracy for the physical comparisons, hence the pattern of results are unlikely to be affected by a speed-accuracy trade-off.

CV_{RT} findings

In terms of CV_{RT} for the semantic comparisons, participants demonstrated more automatic processing of English word-pairs as compared to Malay word-pairs for the first three test sessions. However, on the final test session in day 4, the Malay word-pairs achieved an equivalent level of automaticity as the established English words. In addition the Malay CV_{RTs} were significantly swifter with each test session whereas there was no significant difference in English CV_{RTs}. Therefore, the changes in behavioural effects for the Malay semantic comparisons are more plausibly due to an increase in automaticity of processing the word meanings rather than a general speed-up in RTs due to practice. When considering the time-course of semantic integration, it seems that integration of new words into existing knowledge benefits from extended consolidation as hypothesised in section 4.2. In contrast, there was no significant effect of day on the Malay and English CV_{RTs} for the physical comparisons. The above findings concurred with the hypothesis that there would be no difference in CV_{RTs} with each test session for both the Malay and

English words as participants are making explicit judgments about physical font size rather than semantics.

Overall summary

In summary, findings from Experiment 4 lend further insights into the effects of sleep on new word learning and the time-course for the integration of these new words. The main finding was that the semantic distance effects for newly learnt words were already present at the test session that occurred 15 minutes after training and that there were no further enhancements at subsequent test sessions where participants underwent overnight consolidation. This suggests that the distance effect is more time-dependent and that previous sleep-associated benefits in the distance effects (Experiment 1) might be more related to preserving the new words from further interference. In contrast, the emergence of the size congruity effects for new words in the semantic comparisons is sleep-dependent as the effects only occurred in the test session after participants had experienced a night of consolidation. Moreover, there were differences in the amount of overnight consolidation needed for the sleep-dependent size congruity effects to occur. Findings suggest that a single night of consolidation is sufficient for size congruity effects to emerge in the semantic comparisons whereas for physical comparisons of the new words, there was no main effect of congruity even after four nights of consolidation. Overall, results lend support for both schema learning and system consolidation accounts. Items that are consistent with existing knowledge and relevant to task demands benefit from additional wakeful practice and show time-dependent learning effects. In contrast, for effects that are less relevant to task demands, sleep appears to be key for knowledge integration. Due to the differences in findings for the semantic comparison (intentional automaticity) and physical comparison (autonomous automaticity) tasks, current results predict that when learning new words, greater reliance on sleep-dependent memory consolidation and integration would be needed for processes that require greater task independent automatic semantic access.

Chapter 5: Effects of Sleep and Automaticity on Learning New Numerical Magnitudes

5.1 Introduction

Experiments 1 to 4 have highlighted that sleep and specific sleep components may play an active and causal role towards changes in automaticity in new word learning. This is because sleep was more related to greater semantic distance and size congruity automaticity effects for newly learnt words compared to wakefulness. SWS was also more beneficial for strengthening new form-meaning mappings whereas sleep spindles played a greater role in integrating the new words with existing memories. However, it may also be useful to consider if sleep benefits memory consolidation and knowledge integration of new declarative information beyond the domain of language. For example, there is currently no known research directly investigating the effects of sleep on strengthening and integration in terms of numerical learning, such as learning new numerical-magnitude representations, even though the processing of numerical magnitude is essential in everyday activities and directly related to academic achievement in mathematics (Holloway & Ansari, 2009).

Previous studies have found that knowledge of numerical-magnitude representations can be reflected in terms of automaticity of processing using phenomena like the size congruity effect and numerical distance effect (Ansari, 2008). Therefore, the effects of sleep on new numeral learning can be investigated using similar methods to Experiments 1 to 4, even though the experiments were focused on word learning. This would also enable a more direct comparison of the similarities and differences of the sleep-associated effects on the integration of newly learnt numerals and newly learnt words. The notion that sleep may play a role in new numeral-magnitude learning in addition to new word learning stems from evidence that there may be similar brain areas that represent semantic magnitude information (see Cohen Kadosh, Lammertyn, & Izard, 2008 for a review). The intraparietal sulcus (IPS) that has been widely proposed to implicate the processing of numerical magnitude, has also been suggested to be involved in processing magnitude across different domains (Bueti & Walsh, 2009). In an fMRI study by Cohen Kadosh et al. (2005), the experimenters manipulated Arabic numeral-pairs such that they differed either in numerical distance, physical font or luminance intensity. When asked to make comparisons in each of the domains, the experimenter found increased activation in participants' IPS for the numeral, physical font and luminance domains. Since the above evidence suggests common brain areas for processing magnitude across domains, it may also be possible that the sleep-associated changes when learning new numerals that differ in magnitudes would be similar to the neural changes that occur during memory consolidation of new words that differ in semantic animal size in Experiments 1 to 4.

Even though neuroimaging evidence for common brain areas in processing magnitude information has suggested that similar to new word learning, sleep would also benefit new numerical learning, it is also possible that the time-course for the manifestation of sleep-associated gains would be very different for new numerals compared to new words. Semantic representations for words have also been shown to be different to that of numbers. When considering patients with semantic dementia, a disorder marked by the loss of semantic memory, patients show clear impairments in semantic knowledge such as retrieving meanings of high frequency words but not semantic knowledge in terms of common numerical facts (Crutch & Warrington, 2002; Zamarian, Karner, & Benke, 2006). When listening to words depicting either numerals or body parts, healthy adult participants were significantly quicker to respond to numerals comapred to body parts, even though both sets of words were matched for frequency (Le Clec'H et al., 2000). Moreover, this same result was found when words were presented in the participants' second language and also when words were written instead of spoken. Therefore, it is likely that there will be different rates of learning and subsequent ease of semantic access when comparing newly learnt numerals to newly learnt words from Experiment 1 to 4 due to the cognitive differences between the (representations of) two domains.

5.2 Experiment 5

A recent study by Cohen Kadosh, Soskic, Iuculano, Kanai and Walsh (2010) suggested that automaticity in processing magnitudes of newly learnt numerals may be sleep-associated. In the above study, participants were trained on 9 new numerals that were each associated with an existing numeral ranging from 1 to 9. After training, one of the tests participants underwent a Stroop-like task that was similar to the physical comparison in Experiment 4. Participants were shown pairs of newly learnt numerals (differing in levels of congruity) and had to ignore the numerical magnitude information and select the numeral with the larger physical font size. The participants underwent the above training-test sessions on six occasions across six consecutive days. Even though the main focus of Cohen Kadosh et al.'s (2010) study was on the effect of transcranial direct current stimulation (TDCS) on numerical cognition, there was also a control group in the study that did not receive any TDCS. When looking specifically at the control group, participants only displayed a significant size congruity effect for the new numerals on sessions 5 and 6. As participants were able to experience post-learning sleep between sessions 2 to 6; when comparing Cohen Kadosh et al.'s findings to Experiment 4, it can be implied that sleep promotes increase autonomous automaticity in processing new magnitude information. This is because participants in Cohen Kadosh et al.'s study only displayed significant size congruity effects for the physical comparison task after multiple consolidation nights.

Drawing from the above study, the main aim of Experiment 5 was to investigate the time-course for the emergence of the size congruity and numerical distance automaticity effects for new numeral learning in relation to the moderating role of sleep-associated memory consolidation. The experiment also incorporated sleep polysomnography to investigate if activity during different stages of sleep would be related to different forms of enhancement in automatic access of new numerical magnitude information. Experiment 5 was a close replication of Experiments 1 and 2 where the main difference was that instead of learning 9 new Malay words, participants learnt 9 new numerals that were paired with existing Arabic numerals. After training, participants where either tested immediately, after a 90 minute nap opportunity or after an equivalent period of wakefulness. As comparisons of findings from Experiment 1(overnight study) and Experiment 2 (nap study) have suggested that effects of overnight consolidation was different to consolidation occurring over a 90 minute nap, Experiment 5 also incorporated another group of participants that were tested approximately 24 hours after training and hence experienced overnight sleep. The design in Experiment 5 enabled dissociations between the contributions of time compared to sleep-dependent consolidation toward the enhancements in automaticity for numerical learning. It also enabled the comparison between effects of overnight consolidation and nap-associated effects, which were linked to differences in quality and quantity of sleep. As the longest duration between training and testing was 24 hours, participants in Experiment 5 only had semantic comparisons of the newly learnt numerals. There was no physical comparison task in Experiment 5 as Cohen Kadosh et al's (2010) findings suggested that size congruity effects for newly learn numerals would only emerge in the physical comparison tasks after multiple nights.

Based on Cohen Kadosh et al's (2010) findings, it was predicted that in Experiment 5, sleep would play a beneficial role in the consolidation of the newly learnt numerals where participants who slept after learning would display greater size congruity and numerical distance effects for the newly learnt numerals as compared to those who were tested immediately or remained awake. If sleep is supportive of an active role in numeral learning, then participants who experienced overnight sleep (longer total sleep time and better sleep quality) would also show greater automaticity in processing new numerical magnitude than those in the nap group. It was also hypothesised that participants in the nap and overnight sleep group would display smaller CV_{RTs} (denoting greater automatic processing) for the newly learnt numerals as compared to the participants in the immediate and wake groups. However, as previous research has also highlighted that semantic representations for words were different to that of numbers (Crutch & Warrington, 2002; Zamarian et al., 2006), it was hypothesised that the time-course for the manifestation of gains in automaticity would be different for new numerals in Experiment 5 compared to new words in Experiments 1 to 4.

In addition to the size congruity and numerical distance measures, Experiment 5 also included a speeded mental arithmetic task using the newly learnt numerals. The test was adapted from the one minute addition and one minute subtraction test (Westwood, Harri-Hughes, Lucas, Nolan & Scrymgeour, 1974). Mental arithmetic abilities have often used to be a predictor of early numeral learning abilities and a recent study on 7 year-old children found a positive relationship between quality of numerical magnitude representation and arithmetic performance (Booth & Siegler, 2008). The children with more accurate numerical magnitude estimation in a number line estimation task also displayed greater post-training improvements on new arithmetic (addition) problems. Since sleep has been hypothesised to lead to gains in new numerical magnitude learning, it was predicted that participants who slept after learning would display better performance in the speeded arithmetic test compared to the participants who were tested immediately or those who remained awake between training and testing.

5.2.1 Methods

Apparatus and Materials

Similar to previous experiments, the test apparatus consisted of a computer with a 15-inch screen. The DMDX software (Forster & Forster, 2003) was used to collect all RT and error data. Participants used a USB-joypad for all experimental responses.

Materials also included a speeded number (arithmetic) test for the new numerals. As mentioned in section 5.2, the test was adapted from the one minute addition and one minute subtraction test (Westwood, Harri-Hughes, Lucas, Nolan & Scrymgeour, 1974). For the addition component, all the 30 single digit addition questions from the original test were translated using the novel numerals. For the subtraction component, only 18 questions were translated as 12 of the 30 original questions consisted of two digit numbers. A pilot test revealed that both the addition and subtraction components were matched for difficulty despite having different a number of questions. This was in line with the original study where participants obtained more correct answers for the addition test.

Stimuli and Design

All experimental stimuli and apparatus were similar to Experiment 2 with the key difference that instead of words, the stimuli consisted of nine single-digit Arabic numerals (digits 1 to 9) and nine novel numerals used in Cohen Kadosh et al.'s (2010) study. The training phase consisted of exposure and feedback trials, which were also manipulated in the same way as Experiments 1 and 2 (see Figure 22).

Unlike the training phase, the design of the test phase in Experiment 5 was fairly different compared to all the previous word learning thesis experiments. In test phase of previous experiments, both congruity and distance factors only had two levels (congruent/incongruent and small/large distance). However, in the current experiment, congruity had three levels of manipulation such that the relationship between physical font size and referent size within the numeral-pairs were either congruent (e.g. 1 5), neutral (e.g. 1 5) or incongruent (e.g. 1 5). Hence, the congruent and incongruent numeral pairs differed in both numerical distance and font size, whereas the neutral pairs only differed in numerical distance but not font size. Next, in Experiment 1 to 4, there were only two levels of semantic distances between word-pairs, where word-pairs had either small or large distances. In Experiment 5, there were three levels of numerical distance manipulation where numeral pairs could differ by a distance of 1 (e.g. 1 2), 2 (e.g. 1 3) or 4 (e.g. 1 5). These three levels of numerical distance manipulations were selected based on a previous study that also investigated size congruity and numerical distance effects (O Rubinsten et al., 2002).

All possible pairs of stimuli differing in semantic distance of 1, 2 and 4 were generated, leading to 20 possible item-pairs. Each pair was then also manipulated in terms of congruity (congruent, incongruent and neutral) and counterbalanced on the left and right to prevent biasness in response. In addition, the full set of stimuli was repeated twice (block of presentation). Each participant saw a total of 480 trials: 20 possible distance 1, 2, 4 numeral-pairs x 3 congruity manipulation x 2 (left/right counterbalancing) x 2 block of presentation x 2 (for both existing and new numerals). The total number of trials in Experiment 5 was less than all the new learning experiments where participants in Experiment 1, 2 and 4 saw a total of 576 trials (Malay and English combined) and participants in Experiment 3 saw 518 trials. Still, having three levels of manipulation for congruency and numerical distance increased the level of inter-trial complexity in Experiment 5. The greater inter-trial complexity was thought to have accounted for the reduction in the total number of trials compared to the previous experiments.

A five-way mixed-subjects design was used. The dependent variable was RTs for correct responses. There were five independent variables: one between-subject and four within-subjects. The between-subjects independent variable was the experimental group (nap/wake/overnight/immediate). The within-subjects independent variables were the numeral type (Existing/New numerals), congruity (congruent/incongruent/neutral) and semantic distance (1, 2 or 4) and block of presentation (first or second). Separate analyses were also conducted with proportion of errors as a dependent variable and presented in Appendix 1.5.

Participants

60 monolingual native English speakers were recruited (35 females and 25 males, mean age = 20.5 years, range = 18-27 years) via an online database and were randomly allocated into the nap (n=15), wake (n=15), overnight (n=15) or immediate (n=15) group. All further recruitment criteria and procedure was replicated from the previous experiments.

Procedure

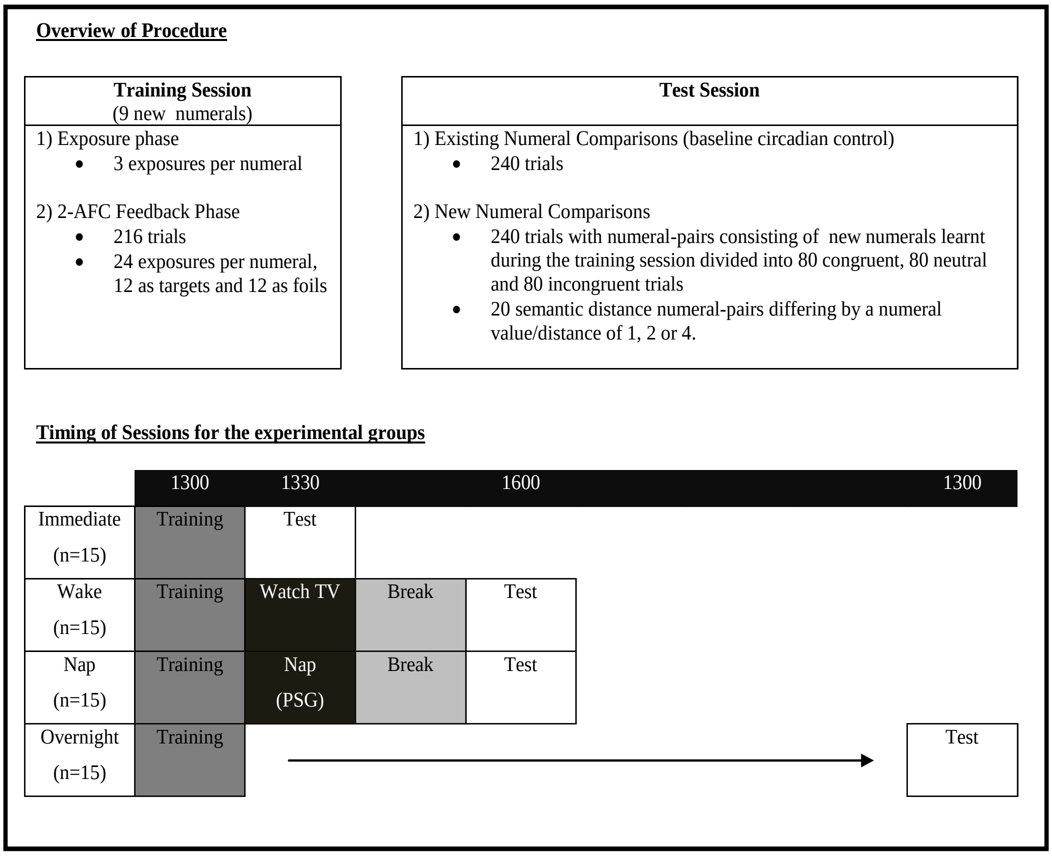


Figure 22. Experiment 5: Summary of experimental procedure including timings of sessions and tasks carried out in the training and test sessions

A summary of the experimental procedure is depicted in Figure 22. All participants arrived at approximately 1:00 pm for the training session. Participants in the immediate group started their test session immediately after the training. In contrast, participants in the overnight group continued with their normal daily activities after the training session and returned to the laboratory for a test session at the same time the following day. The procedure for participants in the nap and wake groups was similar to that of Experiment 2 where participants remained in the laboratory till the test session at 4:00pm. After the training session, participants in the nap group underwent an electrode setup (identical to Experiment 2) for the polysomnographic recording. Between training and subsequent test, participants experienced either a 90 minute nap timed between 2:45 pm and 3:45 pm with sleep polysomnography recording (nap group) or an equivalent period of rested wakefulness (wake group).

The procedure within the training session was replicated from Experiments 1 and 2 with the exception that numerals were used instead of animal names. During the test session, participants were first shown test trials (see stimuli and design section) consisting of existing Arabic numeral-pairs differing in levels of congruity and numerical magnitude distance. They were always instructed to select the larger item in terms of numerical magnitude. Next, participants completed the same comparison task with the novel numerals that they were exposed to during training. After completing all comparison tasks, participants underwent a speeded numerical test (30 addition and 18 subtraction questions) comprising of the novel numerals. Participants were given one minute to complete as many of the addition questions as they could. Participants were then given another minute for the subtraction component. At the end of the whole experiment, the Wide Range Achievement Test (WRAT-3) (Wilkinson, 1993) which measures reading, spelling and arithmetic abilities, was administered.

5.2.2 Results

Data from one participant in the overnight group were excluded as the participant reported disturbed sleep and awakenings for more than 50% of the night. There were no significant differences in Stanford Sleepiness Scale ratings between the four groups at both training and test sessions (all p's > .79). There were also no group differences between WRAT-3 measures reading, spelling and arithmetic abilities (all p's > .30).

Training Session

Table 18. Experiment 5: Mean error rates (in percentages) for the training phase. Standard deviations presented in parentheses.

	Nap	Wake	Overnight	Immediate
Block 1	2.49 (3.40)	2.68 (3.06)	2.47 (2.30)	2.87 (2.41)
Block 2	3.73 (8.38)	2.36 (2.92)	2.50 (2.30)	2.85 (2.26)

Similar to Experiment 2, performance in the training phase was analysed in 2 separate blocks (see Table 18). Participants' error rates were arcsine square-root transformed and analysed in an ANOVA with the experimental group as a between-subjects factor and presentation block as a within-subjects factor. The ANOVA revealed no significant main effects and interactions (all p's > .51).

Test Session

<u>Response Times</u>

Table 19. Experiment 5: Mean reaction times for all correct responses after data trimming. Standard errors are presented in parentheses.

	Existing Numerals				New Numerals			
	Nap	Wake	Over-night	Immediate	Nap	Wake	Over-night	Immediate
Congruity								
Congruent	507 (31)	513 (31)	547 (32)	535 (31)	993 (61)	1138 (61)	1125 (63)	1035 (61)
Neutral	538 (31)	545 (33)	567 (32)	570 (31)	1014 (62)	1152 (62)	1114 (64)	1087 (62)
Incongruent	578 (33)	610 (31)	614 (35)	619 (33)	1043 (60)	1158 (60)	1138 (61)	1122 (62)

Numerical Distance								
1	574 (32)	584 (32)	601 (33)	597 (32)	1065 (62)	1205 (62)	1165 (64)	1143 (62)
2	540 (33)	562 (33)	577 (34)	587 (33)	1034 (58)	1158 (58)	1144 (60)	1096 (58)
4	508 (29)	521 (29)	550 (31)	540 (29)	951 (61)	1083 (61)	1069 (63)	1006 (61)

Unlike previous experiments, RTs from the first and second presentation block were combined as there was no main effect of block for the existing [$F(1, 55) = 1.21$, $p = .27$] and new [$F(1, 55) = 1.31$, $p = .25$] numerals. Mean RTs for all correct responses were separated into six groups based on new/existing numerals and congruent/neutral/incongruent trials. This was done to parallel previous thesis experiments where mean RTs were separated by language and congruity variables. Within each group, trials more than 2 standard deviations away from each participant's mean were discarded and the remaining RTs were presented in Table 19. This resulted in an average data loss of 3.6% and 4.8% per participant for the existing and new numerals respectively.

A five-way mixed ANOVA was performed with experimental group as a between-participant variable and numeral type, congruity, semantic distance and presentation block as within-participant variables. The RT analysis was similar to the previous experiments on language whereby in cases where main effects and interactions were significant for the combined ANOVA with existing and new numerals, further analysis was conducted to investigate if the main effect and interactions were present in both types of numerals or only in the new or existing numerals. Interactions with presentation block were not presented except for cases of theoretical interest.

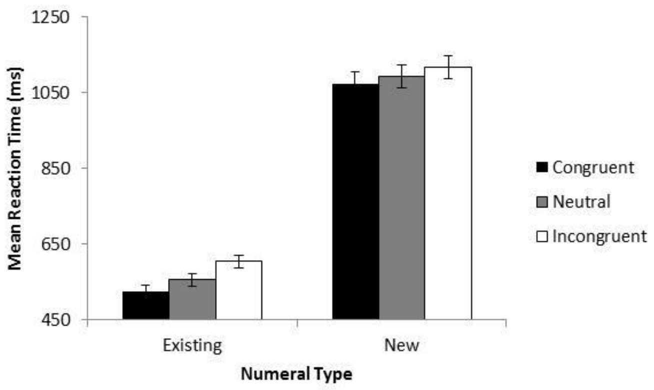
There was a significant main effect of numeral type [$F(1, 55) = 533.99$, $p < .001$] where participants responded significantly quicker to existing numeral-pairs as compared to the new numeral-pairs. In contrast, there was no main effect of experimental group [$F(3, 55) = .66$, $p = .58$] or presentation block [$F(1, 55) = 1.56$, $p = .21$].

In the previous language experiments there were only congruent or incongruent trials, hence only the size congruity effects (RT differences between congruent and incongruent trials) could be investigated. With the inclusion of neutral trials in the current experiment, it was also possible to examine facilitation effects (RT differences between congruent and neutral trials) and also inhibition effects (RT differences between incongruent and neutral trials). There was a significant main effect of congruity [$F(2, 110) = 56.55$, $p < .001$] that was present in both the existing [$F(2, 110) = 149.30$, $p < .001$] and new [$F(2, 110) = 7.96$, $p = .001$] numerals. According to Figure 23a, for both new and existing numeral-pairs, participants had the fastest RTs for congruent trials, followed by neutral trials. Participants displayed the slowest RTs for the incongruent trials.

In addition, there was a significant interaction between congruity and numeral type [$F(2, 110) = 5.37$, $p = .006$] (see Figure 23b). This was because participants displayed a significantly stronger overall size congruity effect [$F(1, 55) = 8.31$, $p = .006$] and inhibition effect [$F(1, 55) = 5.32$, $p = .025$] for the existing compared to new numeral-pairs.[19] In contrast, there was no difference in the strength of the facilitation effects between the existing and new numerals [$F(1, 55) = .98$, $p = .33$]. There were also significant interactions between numeral type, block and congruity [$F(2, 110) = 3.34$, $p = .039$]. Further analysis revealed that the three-way interaction was driven by a two-way interaction between block and congruity that was significant for the existing [$F(2, 110) = 4.21$, $p = .017$] but not for the new numeral-pairs [$F(2, 110) = 1.13$, $p = .33$]. In the case of the existing numerals, the interaction was driven by the neutral trials, which were processed significantly slower in the second compared to first block. There were no equivalent block effects for the congruent and incongruent trials (all p's > .17).

[19] Note: This finding is opposite to the general trend of the animal word data as the English existing words in Experiments 1 to 4 have varying word lengths whereas the Malay words all had fixed word lengths. In contrast both new and existing numerals in Experiment 5 were all singular symbols/digits.

(a)



(b)

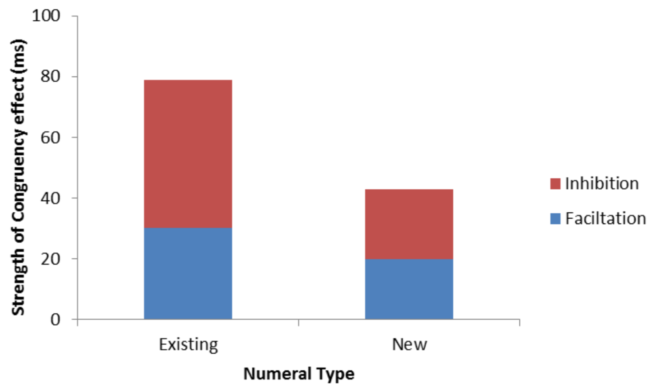


Figure 23. Experiment 5: (a) Mean RTs for congruent, neutral and incongruent trials. Error bars represent standard error of the means. (b) Strength of congruency effects.

Similar to the congruity effect, the semantic distance effect in the previous language experiments only consisted of two levels: small and large. However, in the current numeral experiment, the semantic distance effect consisted of numeral pairs differing by a value of 1, 2 or 4. The main effect semantic distance was significant for the combined ANOVA [$F(2, 110) = 193.25$, $p < .001$] and also significant for both existing [$F(2, 110) = 164.01$, $p < .001$] and new [$F(2, 110) = 92.71$, $p < .001$] numeral-pairs (see Figure 24). There was also a significant interaction between numeral type and distance [$F(2, 110) = 19.30$, $p < .001$]. Participants had significantly quicker RTs for the existing compared to the new numerals for all possible combinations of numeral pairs: when the numeral-pairs differed by a value of 1 [$F(1, 55) = 535.08$, $p < .001$], 2 [$F(1, 55) = 569.11$, $p < .001$] and 4 [$F(1, 55) = 418.86$, $p < .001$].

Figure 24. Experiment 5: Mean RTs for separated by numerical distance within-item-pairs. Error bars represent standard error of the means.

All other interactions in the five-way ANOVA were not significant. It is important to note that unlike the new word experiments, there were no significant interaction between congruity and experimental group [$F(6, 110) = 1.47$, $p = .20$] or numerical distance and experimental group [$F(6, 110) = .53$, $p = .78$] for the newly learnt numerals as respectively depicted in Figures 25 and 26.

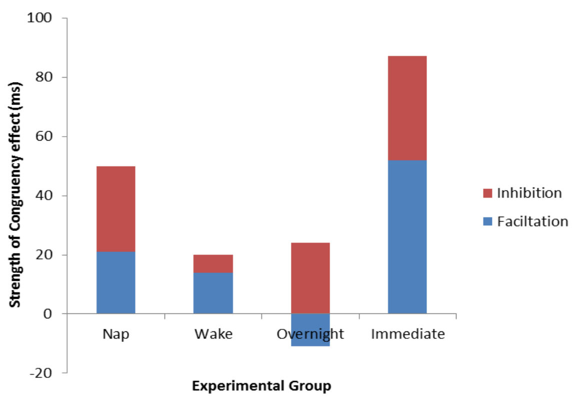


Figure 25. Experiment 5: Strength of congruency effects between experimental groups.

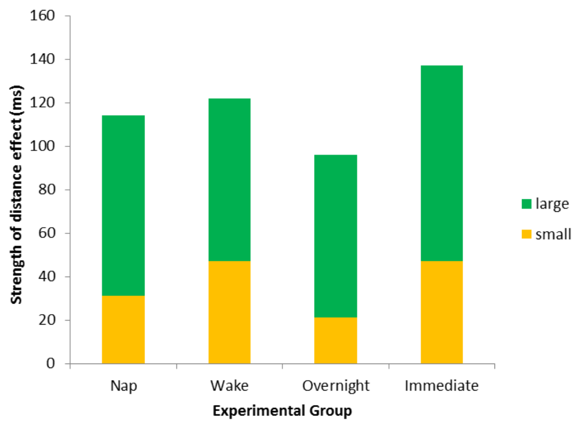


Figure 26. Experiment 5: Strength of numerical distance effects between experimental groups. (small = difference between numerical distance of 1 and 2, large = difference between numerical distance of 2 and 4)

Similar to the previous language experiments, an ANOVA was conducted on participants' CV_{RTS} with numeral type and experimental group as within-participant and between-participant variables respectively, so as to dissociate between a general speed-up in RT and RT changes due to increased automaticity in processing. Similar to the RT analysis, there was a main effect of numeral type whereby participants exhibited smaller (more automatic processing) CV_{RTS} for the existing numerals as compared to the new numerals [$F(1, 55) = 74.32$, $p < .001$]. Unlike the RT analysis, there was also a main effect of experimental group [$F(3, 55) = 3.13$, $p = .033$]. A Tukey's post-hoc tests conducted on all pairwise contrasts indicated that participants in the nap group ($M = .23$, $SE = .016$) had significantly smaller mean CV_{RTS} compared to participants in the immediate group ($M = .30$, $SE = .016$). There were no statistically significant pairwise differences between the means CV_{RTS} for other comparisons between two experimental groups (see Figure 27).

In order to examine if the between (experimental) group CV_{RT} differences were due to circadian confounds, further analyses were conducted separating numeral type. A significant difference in CV_{RTS} between experimental group that occurs for existing numerals would highlight a potential circadian confound whereas an effect for the new but not existing numerals would point towards a benefit of consolidation. When analysing the numeral types separately, results revealed that there was a marginally significant of experimental group for the existing numerals [$F(3, 55) = 2.57$, $p = .063$], whereas, there was a significant main effect of experimental group for the newly learnt numerals [$F(3, 55) = 3.28$, $p = .028$]. This highlights that the main effect of experimental group in the combined ANOVA was likely to be driven by the new numerals. A further Tukey's post-hoc test also revealed that for the new numerals, participants in the nap group ($M = .26$, $SE = .015$) had significantly smaller mean CV_{RTS} compared to participants in the immediate group ($M = .32$, $SE = .015$) (see Figure 27). In contrast, there were no other statistically significant pairwise differences between the experimental groups when comparing CV_{RTS} for the existing numerals.

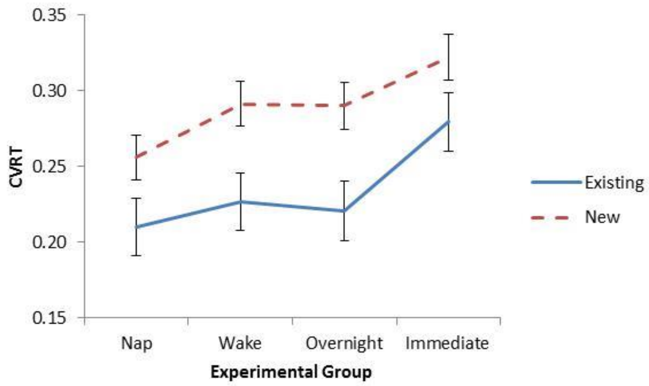


Figure 27. Experiment 5: Mean CV_{RTs} for each experimental group. Error bars represent standard error of the means.

Similar to the previous thesis experiments, further analyses on SDs of participant's mean RT were conducted to examine if the CV_{RTS} results were driven by practice effects or a reduction in RT variance. There was a main effect of numeral type [$F(1, 55) = 383.02$, $p < .001$] but unlike the CV_{RTS} analysis, there no significant main effect of experimental group [$F(3, 55) = 1.76$, $p = .17$]. Therefore CV_{RTS} results relating to differences in experimental group should be interpreted with caution. Still it is not plausible that the CV_{RTS} results are driven by practice effects as there was also no main effect of numeral type for the mean RT analysis.

Speeded Arithmetic Test

Table 20. Experiment 5: Mean performance on speeded arithmetic test. Standard errors are presented in parentheses.

	Nap	Wake	Overnight	Immediate
Raw Score (out of 48)	40 (1.8)	36 (2.5)	43 (1.2)	42 (1.7)
Percentage correct (%)	84.3	74.4	90.1	87.8

The performance for the speeded additional and subtraction tests were analysed together (see Table 20 for untransformed accuracy rates) where accuracy rates were arcsine square-root transformed. A univariate ANOVA was carried out with arcsine square-root transformed accuracy rates as the dependent variable and experimental group as the independent variable. There was a statistically significant difference between experimental groups [$F(3, 55) = 2.96$, $p = .040$]. A Tukey's post-hoc test indicated that participants in the overnight group ($M = 1.32$, $SE = .061$) had significantly better (transformed) accuracy rates than those in the wake group ($M = 1.08$, $SE = .059$). There were no statistically significant differences between all other pairwise group comparisons.

Sleep Stage Analysis

Table 21. Experiment 5: Sleep parameters for participants in the nap group. Parentheses denote the standard error of the means.

Sleep Parameter	Mean time in minutes
Total sleep time	61 (31)
Wake time after sleep onset	19 (25)
Sleep latency	12 (6)
Stage 1	9 (9)
Stage 2	27 (16)
SWS	17 (19)
REM	6 (6)

Note: SWS= slow-wave sleep, REM= Rapid eye movement sleep

Sleep data from 1 participant were excluded from analysis due to a computer error. The main sleep parameters of the remaining 14 participants in the nap group are displayed in Table 21. There were no significant relationships between sleep parameters and behavioural data for participants in the nap group (all p’s > .11).

5.3 Chapter Summary and Discussion

Using the size congruity and semantic distance effects, Experiment 5 explored if sleep would have a beneficial effect towards automaticity in processing new numerical magnitudes. There were two main predictions for Experiment 5. First, it was predicted that similar to the previous new word learning experiments, participants who slept after learning would display stronger size congruity and semantic distance effects for the newly learnt numerals compared to participants who remained awake. However, due to differences in the representations of numerals compared to words, it was also predicted that the time-course for the manifestation of gains in automaticity would be very different for new numerals in Experiment 5 compared to new words in Experiments 1 to 4.

The main findings from the RT analysis suggested that there were no beneficial effects of sleep towards measures of automatic semantic access for the newly learnt numerals. The immediate, wake, nap and overnight group all displayed significant main effects of size congruity and semantic distance for the newly learnt numerals in addition to existing numerals. Moreover, there were no significant group differences for the strength of the size congruity and semantic distance effects for both the new and existing numerals. Therefore, unlike the patterns found in the word learning experiments, participants in Experiment 5 who slept after learning did not display stronger automaticity effects for the newly learnt numerals. The main findings from Experiment 5 suggest that the time-course for the development of automatic semantic access differs between new numerals as compared to new words. It is possible that sleep may not be essential for the consolidation and integration of new numerical magnitude information into existing knowledge.

The current experimental findings are in contrast to that from Cohen Kadosh et al.'s (2010) study that was introduced in section 5.2. Cohen Kadosh et al. (2010) found that participants only displayed significant size congruity effects for their newly learnt numerals on sessions 5 and 6, which occurred after post-learning sleep. Still, one of the key differences between Cohen Kadosh et al.'s (2010) study and the current experiment is that for the current experiment, participants learnt the new numerals similar to a second language learning paradigm. Participants were explicitly taught mappings between the new numerals and existing numerals and hence were likely to be able to rapidly consolidate and integrate the magnitude information related to the new numerals. In contrast, for Cohen Kadosh et al.'s (2010) study, the participants learnt the relationships between the new numbers through feedback from a 2AFC task, where they were not explicitly told the mappings between the new numerals and any magnitude information. Therefore, it is plausible to expect that after the first training session, the new numerals in Cohen Kadosh et al.'s (2010) study were not as well encoded then those in the current experiment. Findings from Experiment 2 (see Chapter 3) also suggested that participants who had weaker encoding performance during training displayed greater sleep (nap)-associated enhancements in automaticity of processing newly learnt words. In addition, as reviewed in Chapter 1, Drosopoulos, Schulze, Fischer and Born (2007) found greater sleep-associated benefits for initially weakly encoded word-pairs than for well encoded word-pairs. This was because following retroactive interference from a new set of word-pair associates, performance on initially weakly encoded word-pairs was better after sleep compared to an equivalent period of wakefulness. These previous findings suggest that the absence of a clear sleep-associated benefit for new numerals in the current experiment may be because the explicit training with existing numerals led to the swift formation of strong numeral-magnitude representations for the new numerals after the training session.

In addition, another key difference between Experiment 5 and Cohen Kadosh et al's (2010) study is that participants in Experiment 5 only made semantic comparisons for the newly learnt numerals whereas participants in Cohen Kadosh et al's (2010) study made physical comparisons. Therefore the size congruity effects in Experiment 5 measure intentional automaticity, unlike the autonomous automaticity measure in Cohen Kadosh et al's (2010) study. As mentioned in the introduction, Experiment 5 only included semantic comparisons of the newly learnt numerals (intentional automaticity) as it was predicted based on Cohen Kadosh et al's (2010) study that autonomous automaticity effects would not emerge after 24 hours. However, it is possible that with the more explicit training in Experiment 5, size congruity effects would emerge after overnight consolidation. If so, findings would be consistent with Figure 21 that was presented in Experiment 4, which predicted that autonomous automatic processing of physical comparisons (of newly learnt items) requires a greater level of sleep dependence than other forms of automatic processing or types of comparisons.

However, the above account may not be sufficient in explaining the difference between the sleep-associated automaticity effects for previous word learning experiments. This is because compared to key findings from the current numerical learning experiment, as both new words and new numerals were taught using a second-language paradigm. The reason for the above differences may be as highlighted in the introduction, where even though there does not seem to be specific brain areas for comparing numerical magnitude, previous research have highlighted clear cognitive differences in the representation of numerals compared to words (Crutch & Warrington, 2002; Le Clec'H et al., 2000; Zamarian et al., 2006). For example, the numeral magnitudes of the digits 1 to 9 may be more fixed and easily quantifiable concepts. In contrast, animal sizes may have more flexible representations in long-term memory. Even though the training for both word learning and numerical learning experiments were similar, the new numerals may have also been more strongly and deeply encoded than the new words. As discussed by Rodd et al.'s (2012) study in Experiment 4, participants had swifter responses towards words that were trained with a deep level of semantic encoding compared to surface encoding even after four days of training. Therefore, unlike the new words in Experiment 1 to 4, the time-course for automaticity effects to emerge in new numerals may be much quicker than the time-course for new words. Drawing back to the schema learning work reviewed in Chapter 1, Tse et al (2007) also mentioned that the more consistent a new schema is with an existing schema, the quicker the newly learnt schema would be integrated into existing knowledge. As the numerical magnitudes are more fixed and easily quantifiable, they may also mirror highly consistent schemas and can be rapidly integrated.

It is also interesting to note that the lack of an interaction between congruity and experimental group suggests that even for the group that was tested immediately after training, participants displayed significant size congruity effects for the new numerals. This is in contrast to Experiment 4 where participants only displayed a significant size congruity effect for the newly learnt words after overnight sleep but not during the test session that occurred 15 minutes after training. Findings in the immediate group in Experiment 5 mirror existing research on rapid integration in schema learning (Tse et al., 2007, 2011; van Kesteren et al., 2012) discussed in the previous chapters. As mentioned in Chapter 1,Tse et al. (2007) demonstrated that rodents which were very well-trained on an experimental flavour-location paired-associated task, were able to learn a new set of flavour-location paired-associates in one trial without the need for any offline consolidation. Moreover, in a recent study, Zhao et al. (2012) taught a group of university students new numerals that were paired with existing magnitude information. During training, the participant had to learn 9 new numerals where each numeral was uniquely paired with a dot array of 10, 20, 30, 40, 50, 60, 70, 80, or 90 dots. Similar to the current experiment, participants were subsequently asked to make semantic comparisons of newly learnt numeral-pairs that varied in numerical distance. The experimenters found a numerical distance effect where participants had swifter RTs for numeral pairs with large compared to small numerical distances. The numerical distance effect was displayed during the test session that was held on the same day as the training session, hence the numerical distance effect emerge before participants had the opportunity for offline consolidation. Similarities in findings between Zhao et al.'s (2012) study, previous schema learning experiments and the current experiment suggests that learning of new numerical magnitude information parallels schema learning accounts were the newly learnt schematic information can be rapidly integrated prior to offline consolidation.

It is also feasible that sleep-associated benefits for new numeral learning may be exhibited in a different way than benefits for new word learning. Despite not displaying sleep-associated gains in measures of automaticity, participants who slept after training showed gains in other measures of learning, such as the speeded arithmetic test where participants were given a maximum time of two minutes to complete 48 addition and subtraction questions with the new numerals. It was predicted that participants who slept after learning would display better performance in the speeded arithmetic test compared to the participants who were tested immediately or those who remained awake between training and testing. Results denoted that participants in the overnight group performed significantly better than those in the wake group for the speeded arithmetic tests. Since gains in arithmetic performance has been previously related to children's accuracy in estimating numerical magnitude (Booth & Siegler, 2008), results suggest that overnight sleep benefited the speed of processing the newly learnt numerical magnitudes as compared to wakefulness. The fact that overnight sleep but not a 90 minute nap led to better performance than wakefulness also suggests that quantity and quality of sleep is important for the consolidation and integration of the newly learnt numerals. This can be paralleled to the word learning experiments where more sleep-associated gains in learning were displayed in the overnight studies (Experiment 1 and 4) compared to the nap studies (Experiment 2 and 3). Still, as there was no significant pair-wise difference in performance between the overnight group and the immediate group who did not experience offline consolidation, it was also possible that extended wakefulness leads to forgetting of the newly learnt numerical information that was avoided if participants were tested immediately. On initial examination, one might be concerned that the results for the speeded mathematics test were confounded by circadian effects. However, there were no significant differences between the experimental groups for the WRAT standardized mathematics subtest. Therefore it is unlikely that the enhanced performance displayed by the overnight group was caused by any circadian effects.

In addition to the speeded arithmetic tests, sleep-associated benefits also emerged for the CV_{RT} measure of automaticity. When analysing the existing and new numerals separately, there was a significant main effect of experimental group for the newly learnt numerals but not for the existing numerals. Post-hoc tests revealed that participants in the nap group had significantly smaller CV_{RTS} when processing the new numeral-pairs compared to participants in the immediate group. Therefore, it can be implied that participants who had an opportunity to consolidate the newly learnt numerals performed with greater overall automaticity in processing during testing as compared to participants who were tested immediately without consolidation. Moreover, the consolidation related gains in automaticity between the post-learning nap and immediate group could be interpreted as sleep-associated as there were no equivalent benefits of a 90-minute period of post-learning wakefulness over being tested immediately. However, it is interesting to note that there was no benefit of CV_{RT} for participants who experienced overnight sleep over those who were tested immediately. One possible reason is that the temporal duration between learning and sleep is crucial. Participants in the nap group slept soon after learning the new words whereas participants in the overnight group remained awake from 1.30pm till their usual bedtime. Even though quantity and quality of sleep is important towards integration of new knowledge, recent studies have also speculated that a shorter temporal duration between learning and subsequent sleep promotes stronger memory consolidation (Hiuyan Lau et al., 2011;

Payne, Chambers, et al., 2012). To further investigate the speculation of whether duration between learning and subsequent sleep affects memory consolidation, subsequent replications can also include an overnight group that is trained in the evening and hence sleeps soon after learning the new numerals. If the duration between learning and sleep/wakefulness is matched between the nap, wake and overnight groups, one can better investigate if quantity and quality of sleep, temporal duration between training and sleep or a combination of both factors is most important in reducing CV_{RTs} (increasing automaticity) for the newly learnt items.

In summary, the current data suggests that numeral magnitude learning seems more related to schema learning as new numerical information may have more fixed magnitude representations that are congruent with existing knowledge. Therefore, the learning and consolidation of new numerical information may be less dependent on sleep especially when participants are only required to make semantic comparisons of the numeral-pairs. It may thus be useful to see if benefits of sleep-associated consolidation would emerge for the size congruity effects if participants were asked to make physical comparisons of the new numeral-pairs. As highlighted in Experiment 4, three consolidation nights were not sufficient for participants to display size congruity effects for physical comparisons of the new words. However, since the learning of new numerals are more 'schema-based', it would be predicted that size congruity effects for the physical comparisons of new numerals would require less sleep-dependent consolidation than new words. It would hence be interesting to replicate Experiment 4 using numerals instead of words and to investigate if the predictions related to Figure 21 would be depicted in the experimental findings.

Apart from the main findings on the size congruity and numerical distance effects, results also highlighted more complex interactions between the levels of congruity or numerical distance and their benefits towards new numeral learning. For the levels of congruity, participants had equivalent performance for the new and existing numeral pairs when comparing facilitation effects (difference between congruent and neutral trials). In contrast, participants had weaker inhibition effects (difference between incongruent and neutral trials) for the new compared to existing numerals. Moreover, existing numerals also had a greater benefit in processing item-pairs with larger numerical distances as compared to the new numerals. This finding is especially interesting in relation to Experiment 1 where sleep-dependent size congruity effects emerged also for trials with large distances (and physical differences). It is possible that similar to Experiment 1, the current findings are driven by differences in saliency within item pairs due to the difference in levels of congruity and numerical distance. An additional avenue for future research would be to investigate if longer term consolidation (beyond a single night) would be sufficient to enhance memory consolidation for the new numerals such that they will display equivalent inhibition effects as existing numerals and also equivalent benefits in automaticity of processing numeral-pairs with larger numerical distances.

In conclusion, unlike the experiments on new word learning, findings from Experiment 5 largely suggest that sleep does not lead to stronger size congruity and numerical distance automaticity effects for semantic comparisons of newly learnt numerals. It is predicted that the limited sleep-associated benefits may be due to the nature of numerical stimuli, whereby numbers represent well-defined schemas with fixed magnitude representation. Hence, newly learnt numerical information can be rapidly integrated prior to sleep. Still, sleep-associated benefits were presented for measures of CV_{RTs} and speeded arithmetic judgments of the new numerals. Therefore, a future replication of Experiment 4 (Chapter 4) using numerals instead of words would be useful, to help investigate the question as to whether the time-course for autonomous automaticity effects for new numbers would be shorter than that for new words and if emergence of autonomous size congruity effects for physical comparisons of new numerals would be sleep-dependent.

Chapter 6: Thesis summary and conclusions

6.1 Thesis summary

This thesis aimed to investigate the effects of sleep on declarative learning and knowledge integration by examining measures of automatic semantic access for meanings of newly learnt items. The aims of the thesis were addressed in five experiments over four chapters. First, Chapter 2 (Experiment 1) addressed how overnight sleep and specific components may play an active role in enhancing automaticity of processing new word meanings. Chapter 3 examined if a shorter 90-minute nap opportunity would lead to changes in automaticity of processing as compared to an equivalent period of wakefulness. Chapter 3 also investigated if between-participant (Experiment 2) and within-participant (Experiment 3) differences in levels of encoding and training would modulate sleep-associated benefits towards new word learning. Chapter 4 (Experiment 4) looked more broadly at the time-course for the emergence of size congruity and semantic distance automaticity effects in new words by investigating the longer term (three consolidation nights) effects of sleep on declarative learning. In the final experimental chapter, Chapter 5 (Experiment 5) extended the research on word learning to explore if the relationship between sleep and new word learning would extend to new numeral learning. The main findings for each chapter will be further summarised below. Broad themes emerging for all experiments, including how findings relate to theoretical models of memory consolidation, novel contributions to literature and areas for future research will then be discussed.

6.2 Main findings

6.2.1 Chapter 2

Experiment 1 used size congruity and semantic distance paradigms to examine whether overnight sleep as compared to an equivalent time awake, would benefit knowledge integration of new word meanings. The size congruity and semantic distance effects are typical of established and well integrated word representations and some researchers also interpret the paradigms in relation to stimulus-response compatibility (Sanders & McCormick, 1976; Kornblum, Hasbroucq & Osman, 1990) where the size congruity effect is thought to require more automatic semantic access than the semantic distance effect as it only emerges with stronger S-R mappings. Participants learnt novel form-meaning mappings where Malay words were taught using a second language learning paradigm. Participants were then tested following overnight sleep or a comparable duration of wakefulness. Participants were also tested on existing English word-pairs that served as a baseline measure to control for fatigue and circadian effects. Previous word learning studies such as those by Dumay and Gaskell (2007) and Tamminen et al. (2010) have highlighted that sleep benefits the consolidation of individual words and the integration of new words with existing lexical knowledge.

With the incorporation of sleep polysomnography, Experiment 1 was also able to explore if specific components during sleep would be related to the size congruity and semantic distance automaticity effects. A relationship between specific sleep parameters and behavioural changes would be supportive of more direct and active effects of sleep on declarative learning. Tamminen et al. (2010) found that sleep spindle activity was positively correlated to measures of lexical integration of the newly learnt words and that SWS was related to the strengthening of individual word memories.

Experiment 1 suggested that sleep benefitted the memory consolidation and knowledge integration of second language word meanings. Participants who slept after learning demonstrated stronger size congruity and semantic distance automaticity effects compared with those who remained awake for Malay word-pairs with large distances and font differences (size congruity effects) and for congruent trials (semantic distance effects). In these particular circumstances, the strength of the semantic distance effect was positively related to SWS and the size congruity effect was positively correlated with sleep spindle activity. Participants in both of wake and sleep groups displayed a main semantic distance effect for the new Malay words. However, only the sleep group displayed a significant main size congruity effect for Malay word-pairs with large distances and font differences.

Despite having baseline English words to control for circadian and fatigue confounds between the sleep and wake group, it was not possible to fully eliminate potential time of day effects in the between-participants design where half the participants learnt the words in the morning and were tested in the evening (wake group) and the half other were trained in the evening and tested the next morning (sleep group).

6.2.2 Chapter 3

Experiment 2 and 3 address the limitations in Experiment 1 by adopting a within-participant nap study design. Even though Experiment 1 adopted baseline measures such as testing participants on existing English words to reduce potential circadian confounds, it was not possible to fully eliminate time of day effects in the between-participants design. By using a within-participants nap study design, it was possible to fully control for time of day effects as participants in the wake and nap sessions were trained and tested at the same timings. Therefore the main difference in design between the wake and nap sessions was the 90 minute nap opportunity and any differences in behavioural performance between the experimental session can be more directly attributed to sleep.

The initial aim of Experiment 2 was to expand on Experiment 1 and investigate if in addition to overnight sleep, whether daytime napping compared to wakefulness would lead to gains in the size congruity and semantic distance effects of new words. This would imply that a shorter period of offline consolidation could lead to benefits in automaticity that reflect measures of knowledge integration. Experiment 2 also used measures of sleep polysomnography to examine if specific sleep components during a shorter period of offline consolidation would also be actively related to changes in behavioural performance. Previous findings have implied that a period of daytime nap that lasts for as short a duration as 6 minutes led to better declarative memory consolidation than wakefulness (Lahl et al., 2008). In addition, the duration of TST, amount of SWS and sleep spindle activity within a nap have been previously shown to be positively related to measures of declarative learning (Alger et al., 2012; Lahl et al., 2008; Schabus et al., 2005).

Results for Experiment 2 indicated a main semantic distance effect for the new Malay words in both the wake and nap sessions, but no size congruity effects in either session. There were also no relationships between the sleep parameters within the 90 minute nap opportunity and any behavioural performance measures. Therefore overall results seemed to suggest that a 90 minute daytime nap was not sufficient to promote measures of knowledge integration in second language learning.

Still, although not part of the initial experimental aim, one of the central findings was that the mean error rate across participants for the whole training phase was close to twice of that in Experiment 1, denoting that there was a large variation in training performance for Experiment 2. When taking into account the individual between-participant differences in training performance, nap-associated benefits in automaticity emerged. For both the semantic distance and size congruity automaticity effects, there were significant relationships between nap-associated gains in strength of the automaticity measures and training performance for the Malay word-pairs. The poorer the participants' training performance, the greater the benefit of the 90 minute nap opportunity over wakefulness on the strength of the semantic distance and size congruity effects exhibited by participants. Close examination of data further highlights that the nap-associated gains were specific to participants who had below average training performance, but not to participants who performed with better than average training performance. Results from Experiment 2 suggested that between-participant variations in training performance were able to modulate sleep-associated memory consolidation and knowledge integration and was supported by previous research looking at between participant variation in encoding strength on sleep-dependent memory consolidation (Drosopoulos et al., 2007).

However there has been limited work examining whether sleep would also selectively benefit automaticity of processing new words that are less well encoded over words that are well encoded within participants. In Experiment 3, encoding strength in new word learning was varied within participants where half the new words were exposed 8 times during training (weak encoding) and the other half exposed for 64 times (strong encoding). It was hypothesised that the within-participant variation in encoding strength would parallel the between-subjects findings in Experiment 2 and Drosopoulos et al's (2007) research where the weakly encoded new words would demonstrate greater sleep-associated benefits in size congruity and semantic distance effects than the well encoded words.

Similar to Experiment 2, participants in Experiment 3 displayed a significant semantic distance effect for the Malay word-pairs in both the wake and nap sessions. In addition, the semantic distance effect was present even for the weakly encoded words. In contrast, there were no main size congruity effects for the new Malay words, with the exception of word-pairs that had strongly encoded targets and foils. With regard to the manipulation of encoding strength, the sleep-associated findings Experiment 3 were less conclusive then that in Experiment 2. The main sleep-associated finding in Experiment 3 was that there was a significant nap-associated benefit towards the strength of the semantic distance effect for Malay word-pairs with strongly encoded foils. There was also a marginally significant nap-associated gain for Malay word-pairs with weakly encoded targets and large semantic distances.

Overall, the role of sleep on within-participant variations in encoding strength was less conclusive than between-participant effects. In addition, there were either significant or marginally significant semantic distance and size congruity effects for the existing English word-pairs in previous experiments except in Experiment 3. This suggests that data from participants in Experiment 3 may be confounded by fatigue effects.

6.2.3 Chapter 4

Experiments 2 and 3 examined the effects of daytime napping and variations in encoding strength on declarative learning and Experiment 1 investigated the effects of overnight sleep quality on declarative memory consolidation. However, there has been limited work exploring longer-term memory consolidation and no known study looking specifically at the time-course for semantic distance and size congruity automaticity effects to emerge for newly learnt items. Hippocampal lesion studies (McGaugh, 2000; Scoville & Milner, 1957) suggest that for many cases, memory consolidation and integration of new memories may sometimes involve an extended time-course. As mentioned in Chapter 4, one key value of experimental studies with multiple consolidation nights is the opportunity to investigate whether sleep has short-term benefits towards memory consolidation (i.e. only after initial acquisition) or if the association between sleep and the time-course of consolidation has long lasting direct effects (Walker & Stickgold, 2010; Winocur et al., 2010).

In Experiment 4, participants were trained on twice as many words (18 words) as the previous experiments. Participants were trained on the words on Day 1 and tested on the same day prior to offline consolidation. Participants were then tested on the words on three consecutive days where they also had experienced overnight sleep. In addition to the semantic comparison tasks used in Experiment 1 to 3, Experiment 4 also included a physical comparison task where participants had to select the word with the larger font size and ignore the meanings of the word-pair. Tzelgov (1999) had previously implied that the size congruity effects in physical comparison tasks were purer and more sensitive autonomous measures of automaticity then the semantic comparisons.

Findings expanded on the time-course for automatic semantic access when making semantic comparisons of new word-pairs. There was a significant semantic distance effect for the Malay word-pairs at the first test session that occurred before overnight sleep. In contrast, the size congruity effect for the first presentation block was only significant after participants had experienced overnight consolidation. Even though it was predicted that after three consecutive consolidation nights, the size congruity effect would be present for physical comparisons of the Malay word-pairs, the findings in Experiment 4 mirrored those in Tham et al's (2011) study where there was no significant size congruity effects for physical comparisons of new Malay word-pairs. The time-course of the semantic and physical comparisons (representing intentional and autonomous automaticity respectively) would be further discussed in section 6.3.2.

6.2.4 Chapter 5

Most systems consolidation models of declarative learning are not specific to new word learning (Diekelmann & Born, 2010a; Frankland & Bontempi, 2005).Therefore, in addition to new second language words, Experiment 5 aimed to investigate the effects of sleep and automaticity in another domains of learning, namely new numeral learning. As the previous thesis experiments have shown differing sleep-associated effects for participants experiencing overnight sleep compared to daytime naps, participants in Experiment 5 were tested 24 hours after training (overnight sleep), after a 90 minute nap opportunity or after 90 minutes of wakefulness. Experiment 5 also included an experimental group that was tested immediately after training.

The time-course for the emergence of size congruity and semantic/numerical distance automaticity effects was shown to be very different in new numerals as compared to new words. Both the numerical distance and size congruity effects were displayed in the new numerals across all experimental groups, even the group that was tested immediately after training. In contrast to the experimental hypothesis, participants who slept after learning did not display stronger numerical distance effects than those who remained awake or those who were tested immediately.

Still sleep-associated benefits emerged for other measures of automaticity and memory consolidation. Arithmetic performance has been previously related to children's accuracy in estimating numerical magnitude (Booth & Siegler, 2008). In the speeded arithmetic test, participants in the overnight group had significantly better performance than those in the wake group. This suggested that overnight sleep leads to better memory consolidation and enhanced processing of newly learnt numerical magnitudes over wakefulness.

6.3 Sleep-associated changes in automaticity for new declarative memory

6.3.1 What is automaticity?

As mentioned in Chapter 4, there is still a debate in the current literature regarding the concept and definition of automaticity (Perfetti, 2007). A strict definition of automaticity includes three main criteria whereby automaticity occurs a) involuntarily (without intention), b) without conscious awareness and c) does not inhibit other courses of action (Carr, 1992; Hasher & Zacks, 1979). However, in the case of word recognition, a broader definition of automaticity is often used (Bargh, 1992; Carr, 1992), where automaticity is defined as an 'unstoppable' process without active effortful regulation that leads to qualitative changes in speed of processing (Bargh, 1992; Segalowitz & Hulstijn, 2005; see Segalowitz, 2003 for a review).

This thesis focused on investigating the effects of sleep on novel word learning using second language paradigms. One of the paramount aims in second language learning is the attainment of 'fluency', which is often characterized by an increase in speed coupled with a reduction in both errors and attentional demands (Segalowitz, 1999; Segalowitz, Segalowitz, & Wood, 1998). The aspects of speed and attention in fluency are also reflected in the definition of 'automaticity' as presented in the previous paragraph. Similar to both sleep and memory, automaticity is not often considered a unitary concept. It is important to point out that the researchers adopting a strict definition of automaticity may not fully agree that the size congruity and semantic distance effects are markers of automaticity. However, both paradigms are considered to satisfy the broader definition of automaticity adopted by researchers exploring the topic of automaticity word recognition and numerical processing (Carr, 1992; Rubinsten & Henik, 2002; Tzelgov et al., 1992).

This thesis first considers differences between the size congruity and semantic distance effects in light of stimulus-response compatibility (SRC) models, which predict that stimulus sets that have relationships consistent with 'real-life' anticipated relationships would result in more effective processing of information (Sanders & McCormick, 1976; Kornblum, Hasbroucq & Osman, 1990). The SRC model predicts that the size congruity effect is related to SRC whereby interference occurs when the mismatch between physical font- referent mappings affects participants response to select the larger (physically or semantically) item. In contrast, the semantic distance effect is not bounded by SRC as the response to select the semantically larger item is not related to the relationship between font-size and referent (animal) size. This highlights that emergence of the size congruity effect requires stronger S-R mappings than the semantic distance effects, suggesting that it is likely that the size congruity effect is a stronger demonstration of automatic semantic access than the semantic distance effect.

In Chapter 4, the thesis also separated the concept of automaticity into autonomous and intentional automaticity as the above division has been widely adopted by scientists who researched on both the size congruity and semantic distance effect (Rubinsten & Henik, 2002; Tzelgov et al., 1992). Automatic processing is considered to be intentional when it occurs as part of a global task requirement. An example of intentional automaticity in word recognition occurs when proficient readers process a sentence, where in order to comprehend the meaning of the full sentence, each word is also individually processed (Tzelgov, 1999). When making semantic comparisons of word-pairs, both the size congruity and semantic distance effect is thought to be related to intentional automaticity as the processing of word meanings within the word-pairs occurs as part of the global task requirement. In contrast, autonomous automaticity is thought to represent a purer measure of automaticity as it involves the influence of variables that are not explicitly task-dependent (see Tzelgov, 1999 for a review). When making physical comparisons of the word-pairs (i.e. to select the larger word in terms of font size), the size congruity effect was thought to reflect autonomous processing as processing the semantic information from the word-pairs was not as explicit task requirement. As autonomous and intentional automaticity reflect different levels of automatic processing in word recognition, it is feasible that both forms of automaticity would require different levels of memory consolidation and knowledge integration. Therefore, one would also expect that there would be different effects of sleep on autonomous automatic processing of newly learnt second language meanings compared to intentional automatic processing.

6.3.2 Time-course and levels of sleep dependence for development of automaticity effects

This thesis provides new insights on the time-course of the development of size congruity and semantic distance automaticity effects in newly learnt second language items and the moderating influence of sleep-dependent consolidation. The main findings from Experiments 1 to 5 are summarised in Figure 28. According to Figure 28, intentional automatic semantic access of newly learnt word and numeral meanings can occur rapidly and prior to offline consolidation. The above evidence was drawn from Experiment 4 and Experiment 5 where both newly learnt second language words and numerals exhibited significant semantic distance immediately after training.

In contrast, the time-course for the development of size congruity effects that require stronger S-R mappings, was different for the words and numerals. In Experiment 5, participants displayed significant size congruity effects for new numerals immediately after training, whereas there was no main effect of size congruity for new word-pairs where participants were tested immediately after training in Experiment 4. Participants demonstrated size congruity effects for nine newly learnt second language words 12-hours after learning if the interval included overnight consolidation (Experiment 1). Participants also showed significant size congruity effects for a larger set of 18 new words when they were tested 24 hours after learning. Hence, the thesis experiments highlight that automaticity effects that require stronger automatic semantic access due to greater S-R mappings would also involve a longer time-course that includes sleep. This is with the notable exception of items that have more fixed magnitude representations in long-term memory, such as numerals.

However, regardless of level of automaticity, participants still display greater knowledge integration measures after sleep for the newly learnt words. Even though the semantic distance effects seemed to emerge independent of sleep, a period of offline consolidation has led to beneficial gains in semantic distance effects compared to an equivalent time awake. For instance participants experienced sleep-associated gains in the strength of the semantic distance effect after a 90-minute nap opportunity, where daytime napping compared to wakefulness led to benefits in intentional automaticity for participants who had poorer training performance (Experiment 2). In Experiment 1, participants who experienced a 12 hour training-test interval that included approximately 8 hours of offline consolidation displayed significantly stronger semantic distance and size congruity effects for newly learnt words compared to participants who remained awake for a comparable duration, regardless of training performance[20]. Therefore, there was a trend where the longer the training-test interval and opportunity for offline consolidation, the stronger the sleep-associated gains in both size congruity and semantic distance effects.

[20] In Experiment 1, an ANCOVA with training performance as a covariate did not reveal any theoretically significant interactions.

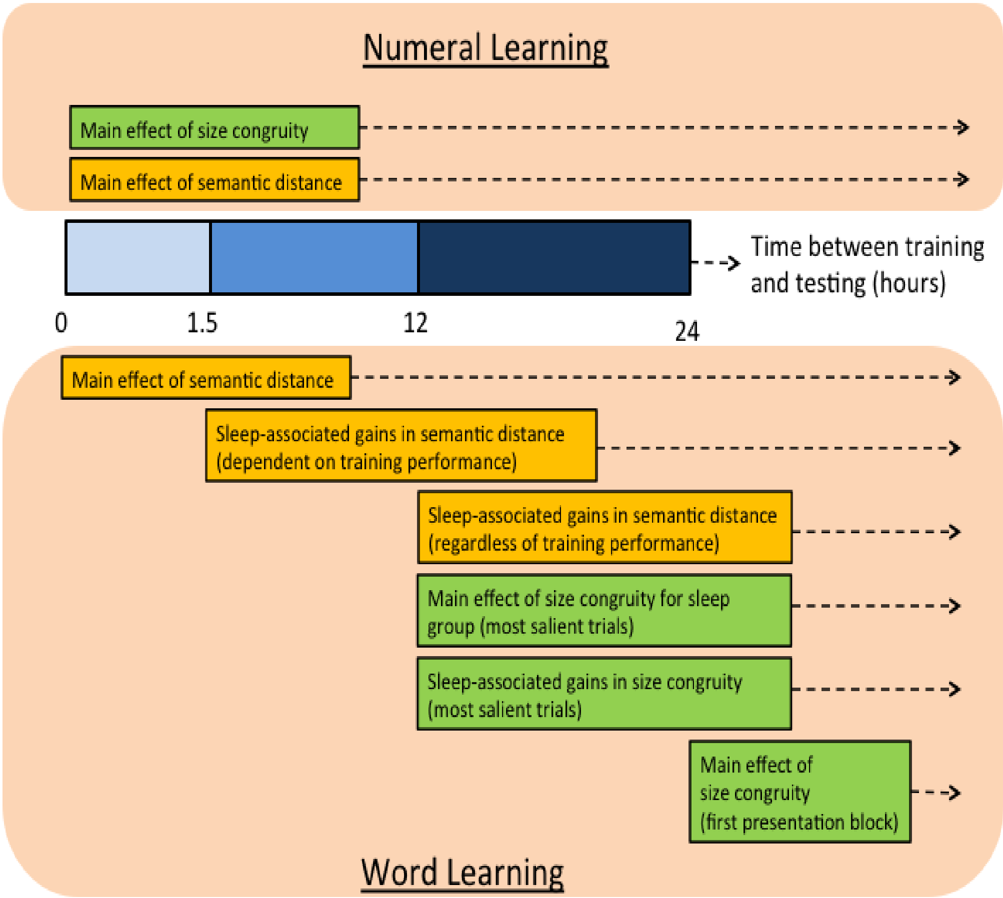


Figure 28. Observed time-course needed for automaticity effects in new numeral learning and new word learning

Still, the results in Figure 28 have only been depicted in semantic comparisons of newly learnt words and numerals, where participants were instructed to select the semantically larger item from the item-pair and ignore discrepancies in font size. Less has been deduced in relation to the time-course for autonomous automatic processing in the physical comparison, where participants had to ignore the meanings of the item-pairs and select the semantically larger item (Rubinsten & Henik, 2002; Tham et al., 2011). Previous work by Tham et al. (2011) suggested that a 12 hour time course that included a minimum of 6 hours' overnight sleep was not sufficient for size congruity effects to emerge in physical comparisons of new words. Experiment 4 expanded the above finding by denoting that a 72 hour time course with approximately 24 hours of sleep was also not sufficient for the development of size congruity effects in physical comparisons.

Although the exact time-course for autonomous automaticity to emerge remains unknown, the thesis experiments provide useful information about the level of sleep dependence needed for semantic distance and size congruity effects in both semantic and physical comparisons of newly learnt second language information. As summarised in Figure 29, sleep dependence was broadly dependent on three factors: the nature of the comparisons, domain of new learnt items and type of automaticity effect.

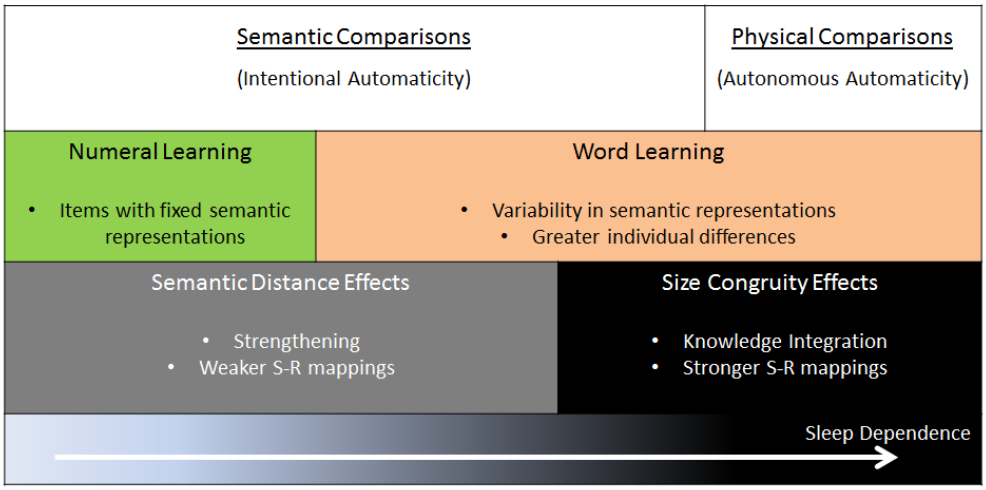


Figure 29. Observed level of sleep dependence required based on comparison task, item domain and nature of automaticity effect

Firstly, greater sleep dependence was needed for autonomous physical comparisons compared to intentional semantic comparisons. Sleep dependence was also related to the domain of the newly learnt items, whereby items with more flexible representations and greater individual differences in representations such as animal names (e.g. a 'dog' could be very large or small in size based on its breed) would be more dependent on sleep for consolidation compared to items with fixed semantic representations (e.g. the meaning of the digit '1'). This topic will be further discussed in section 6.4.3. Similar to the predicted time-course in Figure 28, intentional automaticity effects have been shown to be less sleep dependent that the autonomous automaticity effects. Based on Figure 29, one would predict that size congruity effects for physical comparisons of new items with varied semantic representations would be most dependent on sleep, whereas semantic distance effects for semantic comparisons of new items with fixed representations would be least dependent on sleep for knowledge integration.

6.3.3 Sleep-associated effects on coefficient of variation in RT (CV_{RT})

Table 22. Summary of sleep-associated benefit for CV_{RTs} across experiments

Experiment	**Sleep-associated benefits for CV_{RTs}**	
	New Items	**Existing Items**
<u>Experiment 1(words)</u> 12:12 overnight sleep/wake design (Between participants)	✓ • Sleep group had comparable CV_{RTs} to existing words	✗
<u>Experiment 2 (words)</u> Nap study design - 90-minute nap opportunity - wakefulness (Within participants)	✓ • Only for participants with good training performance	✗
<u>Experiment 3 (words)</u> Nap study design - 90-minute nap opportunity - wakefulness (Within participants)	✗	✗
<u>Experiment 4 (words)</u> Extended consolidation Test 1: 0 hr Test 2: 24 hr Test 3: 48 hr Test 4: 72 hr (Within participants)	✓ • Comparable CV_{RTs} to existing words at 72 hour test session (semantic comparisons)	✗
<u>Experiment 5 (numerals)</u> Four groups based on training-test interval - Immediate (0 hrs) - Nap (90 mins) - Wake (90 mins) - Overnight (24 hrs) (Between participants)	✓ • Nap group had swifter CV_{RTs} than the immediate group	✗

Apart from the size congruity and semantic distance effects, the thesis also investigated effects of sleep on declarative learning and knowledge integration using coefficient of variation in RT (CV_{RT}) measure. A summary of sleep-associated benefit for CV_{RTs} across experiments can be found in Table 22.

The thesis experiments used the CV_{RT} as a marker of the level of automaticity where smaller CV_{RTs} would be a measure of greater automatic access (Segalowitz & Segalowitz, 1993; Segalowitz, Segalowitz, & Wood, 1998). The experiments also considered the difference between CV_{RTs} for the new learnt items compared to CV_{RTs} for the existing well-integrated items. Lack of a difference between CV_{RTs} for the new and existing items was also considered a marker of enhanced integration as that would denote that consistency and stability of responses to new items were similar to well-integrated items.

Participants in the sleep group had comparable CV_{RTs} between new and existing words after a night of consolidation (Experiment 1). When participants were trained on twice as many words (Experiment 4), sleep-associated benefits for CV_{RT} emerged after three consolidation nights. In contrast, there were no sleep-associated benefits for CV_{RT} across all participants when they only experienced a 90 minute nap opportunity. When participants experienced a shorter duration of sleep (Experiment 2), sleep selectively enhanced CV_{RTs} for participants who had exhibited good training performance. Therefore, the main finding across the thesis experiments is that the duration of offline consolidation was important for enhancement in CV_{RTs} and that for shorter durations of consolidation, sleep plays a selective role towards benefitting CV_{RTs} for participants who had good training performance.

It should be noted that even though the CV_{RT} measure has been fairly widely adopted in the second language literature to distinguish between increased automaticity due to language proficiency (reduction in CV_{RT}) as compared to a practice related speedup in reaction time (reduction only in mean RT), there has also been strong limitations in previous CV_{RT} experiments that prevent researchers from drawing conclusive findings (Cox & Calderon, 2014). Key limitations in current literature include the lack on an independent experimental measure of baseline second language proficiency. For example, the findings reported often use baseline measures reported by individual schools that use different criteria assess proficiency, which makes the findings between studies less comparable (Cox &

Calderon, 2014). Still as compared to changes in participant's standard deviations, the CV_{RT} may still be a more reliable measure of automaticity of processing as it is more independent from practice-related mean RT speedup. Previous second language learning studies have often found strong correlations between changes in standard deviations and changes in mean RTs, whereas there were only weak correlations between CV_{RTs} and mean RTs (Segalowitz & Segalowitz, 1993; Flehmig, Steinborn, Langner, Scholz, & Westhoff, 2007).

This thesis provides the first known research using CV_{RTs} to examine the effects of sleep on automaticity and memory consolidation in second language learning. The thesis also considers sleep-associated changes in CV_{RTs} beyond the language domain. Although research on CV_{RTs} has focused on second language learning, Experiment 5 also suggests that sleep may benefit CV_{RTs} for new numeral learning. Still the sleep-associated effects for new numerals seem less conclusive and more research would have to be done to validate in sleep would also enhance CV_{RTs} beyond the word learning.

6.3.4 Insights to the concept of automaticity in declarative learning

This thesis adopted a two-pronged approach when investigating sleep-associated changes in automaticity. Drawing from word recognition and numerical magnitude processing studies, the thesis first considered the intentional and autonomous aspects of automaticity using the semantic distance (intentional) and size congruity (autonomous) paradigms (Rubinsten et al., 2002; Rubinsten & Henik, 2002; Tzelgov et al., 1992). The thesis also used the CV_{RT} as a marker of automaticity. In contrast to a reduction in RTs, which signifies a quantitative change in performance using the same 'controlled' processing mechanism, a reduction in CV_{RTs} marks swifter performance due to a reduction or removal in the previous 'controlled' processes (Segalowitz & Segalowitz, 1993).

Instead of considering the intentional and autonomous concept of automaticity as a conflicting view to the CV_{RTs} qualitative reduction in RT variation view of automaticity, this thesis suggests that both interpretations of automaticity can be integrated to gain new insights on the concept of automaticity in declarative learning. Both the size congruity and semantic distance paradigm and CV_{RT} measure view automaticity as a change in speed of processing. Changes in speed of processing can be due to practice effects or training effects that lead to a quantitative reduction in RT. However, results from the thesis highlights that changes in automaticity are marked by qualitative changes in RTs that may or may not be coupled by a quantitative change.

As mentioned in the previous paragraph, automaticity can be broadly depicted as a qualitative change in speed of processing. Within the domain of the qualitative change, levels of automaticity can be depicted as a continuum with some processes 'more automatic' then others. Findings from this thesis suggest that for new declarative learning, quantitative changes in speed of processing are not sleep-dependent. As mentioned previously, factors such as practice, training and wakeful passages of time seem to be sufficient for quantitative changes in speed of processing (Segalowitz & Segalowitz, 1993; Stickgold & Walker, 2007). Unlike quantitative changes, benefits to automaticity in terms of overall qualitative change in speed of processing, are sleep-associated. The thesis also suggests individual differences in sleep lead to differences in the level of automatic processing achieved based on the 'automaticity continuum'. For example, in Experiment 1, participants in the sleep group experienced a greater reduction in CV_{RTs} than participants in the wake group (overall qualitative change). However, within the sleep group, individual differences in sleep also lead to further differences in automatic processing, whereby participants with greater amounts of SWS and greater sleep spindle activity displayed stronger semantic distance and size congruity effects respectively (specific sleep-associated differences reflected on automaticity continuum).

6.4 Contributions to and relationships with existing theoretical models

6.4.1 Dual Process Hypothesis

As mentioned in section 1.6.1, the dual process hypothesis proposes that SWS would benefit the consolidation of declarative memories whereas REM sleep would benefit procedural memories (Peigneux et al., 2001). Findings from this thesis support the dual process hypothesis to a small extent. Results from Experiment 1 highlighted that participants with greater SWS also displayed stronger semantic distance effects for the newly learnt word-pairs. Still, in addition to SWS, Experiment 1 also highlights that sleep spindle activity is positively related to knowledge integration, where there was a positive correlation between spindle activity and strength of the size congruity effects for the Malay word-pairs. Even though sleep spindles are present in SWS, they are mainly found in stage 2 sleep (Rechtschaffen & Kales, 1968). Therefore this thesis suggests that the dual process hypothesis is not a sufficient model for explaining sleep-dependent knowledge integration as it does not account for the contribution of other sleep stages or sleep components that have been found to be related to knowledge integration.

6.4.2 Sequential Hypothesis

In contrast to the dual process hypothesis, the sequential hypothesis model predicts that memory consolidation is driven by the complementary and sequential role of NREM-REM sleep cycles (Giuditta et al., 1995). It is not feasible to make direct comparisons between current thesis findings and the sequential hypothesis model as the current data analysis did not examine the role of sleep cycles on measures of knowledge integration.

An interesting avenue for future research may be to consider if the sequential hypothesis model can explain the differences in findings between the overnight and nap studies. Sleep-associated gains in size congruity effects for newly learnt words were only present after overnight sleep but not after a nap. Participants also experienced greater benefits towards CV_{RTs} of the newly learnt words after overnight sleep compared to after daytime nap. If knowledge integration is related to the sequential role of NREM-REM sleep cycles, then it can be predicted that when controlling for total sleep time and time spent in each sleep stage, participants who experience overnight sleep after learning would still display greater levels of knowledge integration than those who experienced a nap. This is because a period of overnight sleep consists of more NREM-REM sleep cycles than a 90 minute nap. For example, it was indicated in section 6.3.3 that the duration of offline consolidation was important for enhancement in CV_{RTs} where greater sleep-associated benefits for CV_{RTs} were exhibited after longer durations of consolidation than shorter durations of consolidation. It would be useful to further examine the above relationship to see if the increased number of NREM-REM sleep cycles is a more crucial factor than sleep duration in influencing CV_{RTs}.

6.4.3 General Two-stage Systems Consolidation and Complementary Learning Systems

The general two-stage system consolidation model proposes that memory consolidation involves two main systems: the hippocampal and neocortical system. The hippocampal system acts as a short-term store for new and distinct declarative memories, whereas the neocortical system serves as a permanent store for long-term memories (Diekelmann & Born, 2010a; Frankland & Bontempi, 2005).

The complementary learning systems (CLS) model draws upon the two-stage system consolidation model. According to the CLS model (McClelland, McNaughton, & O'Reilly, 1995) of word learning (Davis & Gaskell, 2009), new declarative memories are encoded in both hippocampal and neocortical systems through interconnected networks, but rely more heavily on the hippocampal system during initial learning. However, when the brain is in an offline state such as sleep, the new memory representations are reactivated in the hippocampal system and increased reliance is placed on connections in the neocortical networks. After multiple cycles of offline reactivation, memory representations are integrated into existing networks in the neocortical system and eventually become independent of the hippocampal system during retrieval (Diekelmann & Born, 2010a).

McClelland (2013) also expanded on the CLS model by suggesting that when newly learnt items are highly consistent with existing knowledge, rapid hippocampally mediated transfer of information to neocortical system occurs. The finding that semantic distance effects were displayed for new Malay word pairs immediately after learning (Experiment 4) supports McClelland’s (2013) account. This is depicted in Figure 30a, where mappings between the new Malay word forms and existing information about animal sizes are formed after learning and prior to offline consolidation. The rapid hippocampally mediated emergence of form-meaning Malay mappings occur in the thesis experiments as the new words are taught as a second language, hence are consistent with and relevant to existing linguistic knowledge.

The thesis findings add new insights on how specific components of sleep might operate in the systems consolidation and CLS model of word learning. Experiment 1 provides some of the first evidence in literature where for novel word learning, SWS and sleep spindles benefit size congruity effect and semantic distance effect processing, respectively. Experiment 1 suggested that SWS may play a main role in strengthening individual episodic form-meaning mappings in the neocortex, whereby reactivation of hippocampal memories during SWS enables the individual mapping to be more reliant on the neocortical system for storage and retrieval (see Figure 30b). As the size congruity effect involves the influence strong S-R mappings, an increased level of knowledge integration may be needed for the effect to emerge. It can be speculated from the results in Experiment 1 that sleep spindle activity and their interplay with slow oscillations and sharp-wave ripples may be more crucial for the integration of the newly learnt form meaning mappings with existing knowledge about animal size relationships (see Figure 30c). It is predicted that the combined role of SWS and sleep spindles enabled the formation of new neocortical representations of individual Malay words and integrated connection between words, which then led to generalised associations between the new representations (see Figure 30d).

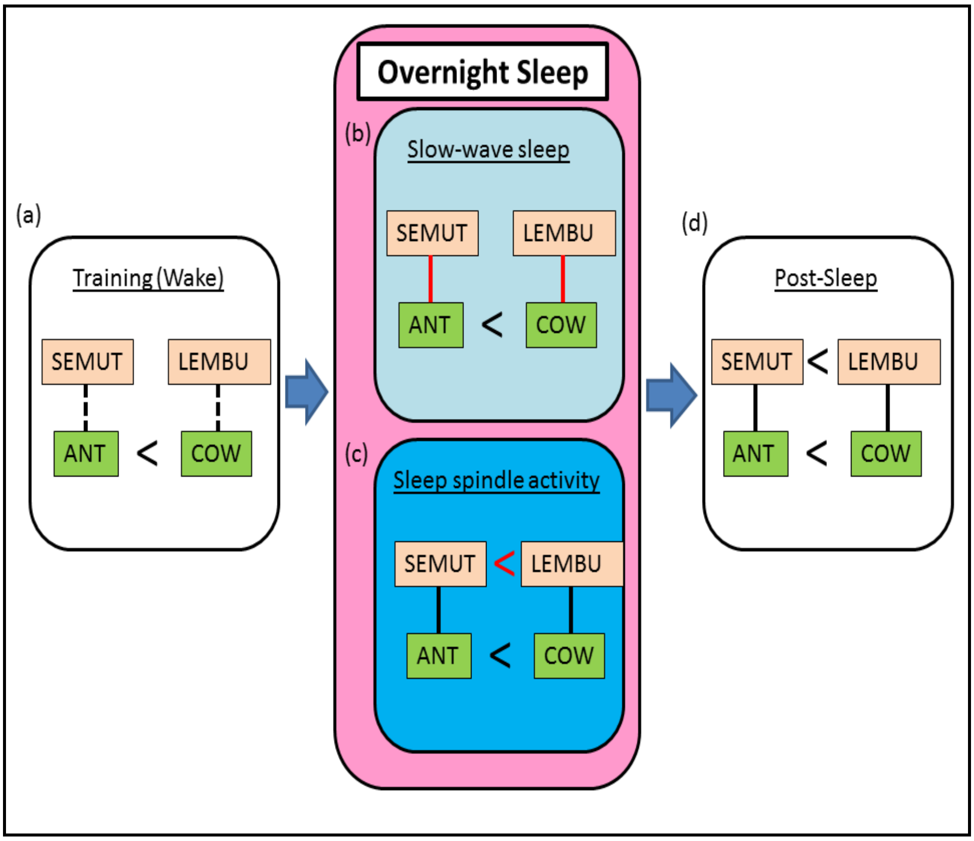


Figure 30. Contributions of SWS and sleep spindle activity on automatic processing of new word meanings. The connecting lines depict the strength of the relationship between mappings: the solid lines denote a stronger relationship than the dotted lines. The arrows highlight the knowledge of the magnitude representations between different animals. The lines and arrow in red denote the predicted active sleep-associated changes. (a) After training, mappings with new word forms and existing semantic representation emerge for the Malay words learnt as a second language. (b) Offline reactivation of the new memories during SWS enables further strengthening of the newly learnt form-meaning mappings . (c) Interplay between sleep spindles, slow oscillations and sharp-wave ripples help integrate the new form-meaning mappings with existing knowledge (about animal size relationships). (d) The combined role of SWS and sleep spindles led to post-sleep gains in strength of size congruity and semantic distance automaticity for the new words

Although there were no relationships between sleep parameters and automaticity effects when adopting a nap design, Experiments 2 and 3 still contribute to the CLS model of word learning. Both experiments predict that a minimum duration of offline consolidation is required to play an active role in memory consolidation and knowledge integration. For second language learning, a 90-minute consolidation opportunity was not sufficient to promote increased semantic access to new word meanings in cases where greater S-R mappings were required whereas overnight sleep (approximately 8 hours) was sufficient in benefitting the development of measures of knowledge integration such that size congruity effects were present for the newly learnt words. It can be predicted according to Figure 30c that the differences in sleep-associated gains in size congruity effects were due to differences in spindle activity between a daytime nap and overnight sleep. Another case where there was no significant sleep enhancement in behavioural performance was the autonomous physical comparisons in Experiment 4. It has been predicted that three nights of offline consolidation was not adequate for the level of strengthening (Figure 30b) and integration (Figure 30c) needed for size congruity effects to emerge in the physical comparisons. Therefore, it can also be speculated that after longer term consolidation with increased strengthening of form-meaning mappings and increased integration with existing knowledge, the size congruity effect will emerge for physical comparisons of new words.

According to the CLS model of word learning (Davis & Gaskell, 2009), sleep-associated differences in knowledge integration in relation to RTs for lexical competition effect (Dumay & Gaskell, 2007; Tamminen et al., 2010) reflect rapid access to the word form as part of the lexical competition process. While before sleep participants can access the novel forms (recognition memory), the access is not fast or automatic enough to influence lexical competition process. However, after sleep, automatic access of the novel word forms is displayed through the lexical competition process. In addition to automatic access of novel word forms, the degree of integration of new linguistic knowledge can be determined by the extent to which newly learnt words exhibit properties of existing words in the lexicon (Davis & Gaskell, 2009), such as rapid and automatic access of word meaning.

Besides findings on the size congruity and semantic distance paradigms, the thesis also expands on the CLS model of word learning using findings from the CV_{RT} measure. As mentioned in the previous paragraph, the lexical competition effect reflects rapid access to the word form. As the CV_{RT} measures variation in standard deviation rather than absolute RT, sleep-associated reduction in CV_{RTs} may be interpreted to reflect consistency and stability in the access of word form and meaning (Segalowitz et al., 1998). The thesis findings regarding sleep-associated effects on CV_{RTs} (depicted in section 6.3.3) contribute to the CLS model of word learning by providing evidence that sleep not only leads to rapid access to newly learnt word forms (lexical competition) and more automatic access of word meaning (size congruity and semantic distance effects), it also enables more consistent and stable processing of word forms and meanings.

6.4.4 Synaptic Homeostasis

Unlike the systems consolidation and CLS models that attribute the memory consolidation and knowledge integration process to the active offline reactivation during sleep, the synaptic homeostasis model highlights that memory consolidation is the consequence of synaptic downscaling that occurs during sleep (Diekelmann & Born, 2010b). The synaptic homeostasis model also implies that slow-wave activity (SWA) is particularly important for synaptic downscaling through the pruning of synaptic weight (Tononi & Cirelli, 2006). As section 1.6.5 mentioned, there are limited insights in current literature of how the synaptic homeostasis model relates to knowledge integration.

As SWA is a marker of SWS, Experiment 1 suggests that the positive relationship between SWS and the semantic distance effect for Malay words may have been a product of a consolidation and integration process that occurred during synaptic downscaling. Moreover, the synaptic homeostasis also suggests that wakefulness may not be ideal for memory consolidation due to the synaptic potentiation process that leads to a saturation in synaptic strength (Tononi & Cirelli, 2003). The saturation in synaptic strength and the lack of synaptic downscaling during wakefulness may be a useful explanation as to why results from the thesis experiments have indicated that an equivalent duration of wakefulness does not lead to comparable benefits to sleep for knowledge integration of second language words.

It should be noted that a comprehensive overview of mechanisms within the synaptic homeostasis and current experimental findings is beyond the scope of this thesis. Overall, it is very likely that the systems consolidation, CLS and synaptic homeostasis models all relate to different aspects of sleep-associated knowledge integration (Diekelmann & Born, 2010a). Still, a specific point to note is that there are currently no predictions within the synaptic homeostasis model regarding the role of sleep spindles of synaptic downscaling and memory consolidation (Tononi & Cirelli, 2006). In light of the relationships between sleep spindle activity and the size congruity effects in Experiment 1, the relationship between sleep and knowledge integration in this thesis may be more related to the general systems consolidation and CLS models.

6.5 Selective factors that modulate sleep-associated consolidation

One of the aims of this thesis was to examine whether sleep-dependent consolidation of declarative memory would be modulated by between-participant and within-participant manipulations of encoding strength and saliency, to better understand the implications of sleep on declarative memory and learning.

6.5.1 Between-participant individual learning ability

As mentioned in the Chapter 1 literature review, between-participant differences in the level of encoding during learning (or training) is one of the selective factors that may modulate sleep-associated consolidation. Existing research has indicated conflicting findings regarding the direction of the modulating effect. Drosopoulos, Schulze, Fischer and Born (2007) implied that sleep selectively enhances performance in participants who formed weakly encoded presentations of the newly learnt items as compared to participants who had form strongly encoded presentations. In contrast, Tucker and Fishbein (2008) found the opposite relationship whereby daytime napping benefitted word-recall performance of participants who performed well during training but not those who performed poorly during training.

If training performance were viewed as a measure of encoding level, whereby poorer training performance would be linked to weaker encoding of new items, then findings from Experiment 2 would provide more support to Drosopoulos et al.’s (2007) study as nap-associated enhancements in performance selective to participants who performed poorly but not strongly during training. Still, the between-participants differences in encoding were explicitly manipulated in Drosopoulos et al.’s (2007) study, where participants who had stronger encoding were given more feedback. The between-participant differences in Experiment 2 were more closely linked to Tucker and Fishbein’s (2008) study where poor and good training performance was not due to explicit experimental manipulation but rather individual differences in participants' learning abilities.

The contradictory findings on the benefits of sleep for strong encoding (good learning ability) and weak encoding (poor learning ability), seem to suggest that sleep may not play a selective role in benefitting either strongly or weakly encoded items. Stickgold (2009) predicted that moderately encoded items would experience the most sleep-dependent gains. Items that are very weakly encoded may not be resistant to interference or forgetting prior to the sleep opportunity and items that are very strongly encoded may not need sleep to actively strengthen and integrate the new memory representations. Figure 31 in section 6.5.2 suggests how the discussed findings related to Stickgold's (2009) predictions.

It is important to note that encoding strength is subjective to a certain extent and it highly plausible that the measure of the strength of encoding in Tucker & Fishbein's experiment is very different to that of Experiment 2 (current thesis). Figure 31 suggests that the weak and strong encoding in Tucker & Fishbein's experiment may mirror, respectively, the very weak and moderately weak encoding in Stickgold's (2009) predictions. In contrast, the weak and strong encoding in Experiment 2 may be mirror, respectively, the moderately strong and very strong encoding in Stickgold's (2009) predictions. The above differences would explain the different trends in sleep selectively for Tucker & Fishbein's experiment and Experiment 2 in the current thesis.

Still, the effect of between-participant differences in the level of encoding during training was not replicated in the other thesis experiments. The lack of replication of findings could be because the other experiments were overnight studies[21]. It is plausible that the selective effects of sleep on the above between-participant difference only occurs when there is a short opportunity (duration) for consolidation. In contrast, with a longer opportunity for consolidation (overnight sleep), there may be less selective effects of sleep on memory consolidation. Therefore, before proceeding to further investigation of between-participants factors that modulate sleep-associated consolidation, it is important to replicate the findings from Experiment 2 to see if the results found are reliable and generalizable.

---

[21] It was not possible to conduct a direct comparison on generalisability between Experiment 2 and 3 as the procedure for the training session in Experiment 3 was very different (e.g. the training trials were manipulated for encoding strength).

6.5.2 Stimuli specific effects

Apart from between-participant individual differences in learning, stimuli specific effects were also predicted to moderate the selectivity of sleep-dependent consolidation. According the literature review in Chapter 1, existing research proposes that sleep selectively benefits the consolidation of items that are more salient over items that are less salient (Hu et al., 2006; Payne et al., 2012). However the review also highlighted that research on the effects of item saliency on sleep-dependent consolidation has largely been limited to emotional memory consolidation.

Experiment 1 suggests that item saliency also affects the selectivity of sleep-associated consolidation effects on word learning whereby the sleep-dependent benefit in size congruity effect was only present for Malay word-pairs with larger physical differences. Adopting the definition of saliency as a factor that causes increased cognitive awareness such that the item or thought stands out (Myers & Alpert, 1977), saliency may be related to disparity in font size as word-pairs with larger font size disparities are more visually distinct. Beyond visual saliency, items with larger semantic distances may also be more salient when considering the mental imagery whereby participants' mental imagery of item-pairs with large distances would be more striking (Paivio, 1975). Out of all the possible manipulations, items with large semantic distances and large physical are most salient visually and mentally and hence the current findings concur with existing literature, and also implies saliency can affect memory consolidation in word learning.

In addition to saliency, stimuli specific effects may also refer to encoding strength. As mentioned in section 6.5.1, encoding strength may be related to between-participant individual learning abilities where good learner form more strongly encoded presentations of newly learnt items then poor learners. It is also possible to investigate within participant differences in strength of encoding, for example in Experiment 3 where half the new words were weakly encoded and the other half were strongly encoded. Unlike Experiment 2, there was no clear sleep-associated gain for either the weak or strongly encoded items. It is plausible the weak and strong encoding in Experiment 3 map onto the extreme ends of Stickgold's (2009) predictions where neither extremes show sleep-dependent consolidation benefits (see Figure 31). In order to further validate Stickgold's

(2009) predictions, it would be useful to replicate Experiment 3 to include items that are encoded to a moderately weak and moderately strong level and test if those items would experience sleep-associated gains in size congruity and semantic distance effects.

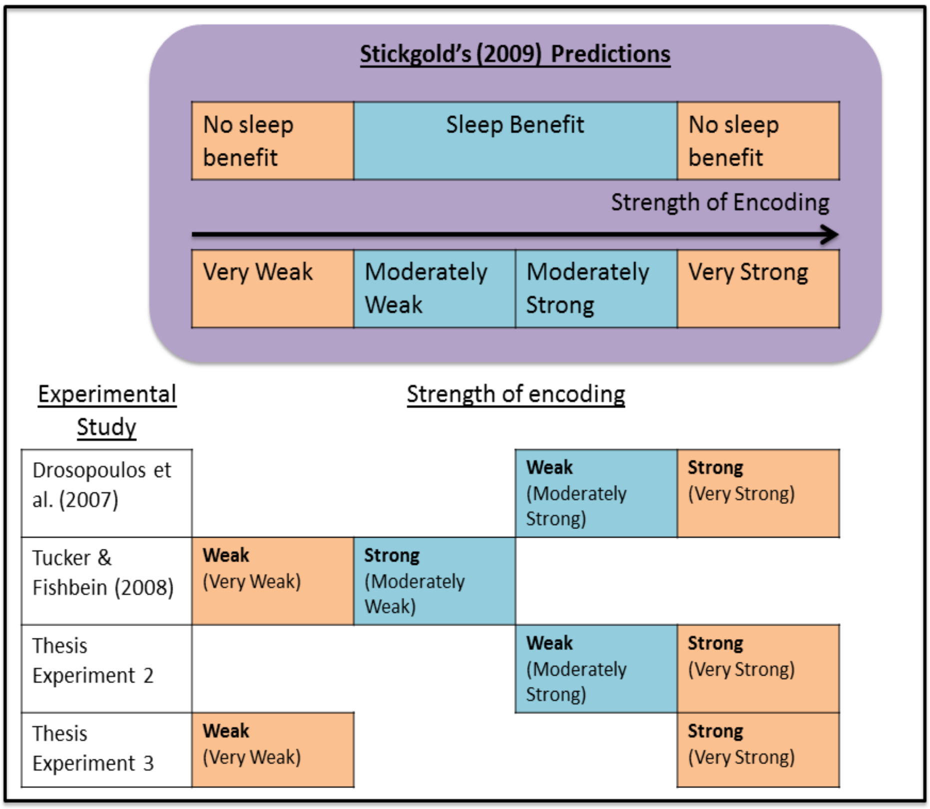


Figure 31. Strength of encoding and sleep-associated benefits in relation to Stickgold's (2009) predictions (in the purple area). Individual boxes in blue are predicted to experience sleep-associated benefits. Boxes in orange are predicted not to experience sleep-associated benefits. The words in bold denote the level of encoding strength depicted by the authors of the individuals studies. In contrast, the words in brackets are interpretations of encoding strength relative to Stickgold's (2009) predictions.

The sleep-associated effects for Experiment 3 also seem more related to ease of processing rather than encoding strength. There was a significant nap-associated benefit towards the strength of the semantic distance effect for Malay word-pairs with strongly encoded foils. It is predicted that word-pairs with strongly encoded foils would be more difficult to process then word-pairs with weakly encoded foils. As mentioned in Chapter 3, an area for future research may be to manipulate encoding difficulty of the new words instead of encoding strength. It would be interesting to see whether items that participants felt were more difficult to encode would experience greater selective sleep-associated enhancements compared to items that participants felt were easy to encode.

6.5.3 Domain specific effects

Section 6.3.2 predicted that level of sleep dependence towards memory consolidation was related to the domain of the newly learnt items. In general, learning new numerals seem to require a smaller level of sleep dependence than new words. It was proposed that this difference was due to how fixed the ‘first language’ representations were in long term memory (see section 6.3.2).

When examining Figure 29, one might notice that there were no observations for physical comparisons of new numerals. This was mainly because in Experiment 5, only the semantic comparisons of the newly learnt numerals were tested. It would be useful to also replicate Experiment 5 where physical comparisons are tested as this will add insights as to whether item domain or the nature of the comparisons during testing is a more important factor for moderating sleep effects. If the nature of the comparisons are more dependent on sleep, then emergence of automaticity effects for physical comparisons of new numerals would take a shorter time-course than that for physical comparisons of new words, but would require a longer time-course than semantic comparisons of new words. In contrast, if the item domain is a stronger moderator of sleep selectivity, then automaticity effects for new numerals would emerge prior to that for new words regardless of the nature of the comparisons made.

6.6 Weakness of the present thesis

6.6.1 Effects of sleep inertia

A possible weakness of this thesis is the potential effects of sleep inertia for the sleep and nap groups. Sleep inertia can be broadly defined as an occurrence that happens immediately after sleep where for a period, post-sleep performance is significantly worse than pre-sleep performance due to a loss of alertness (Achermann & Werth, 1995). A review by Tassi and Muzet (2000) highlighted that current research does not suggest a conclusive time-course of sleep inertia. Majority of the studies reviewed suggest that sleep inertia does not less beyond 30 minutes (Tassi & Muzet, 2000). Therefore, all participants in the sleep and nap groups (or sessions) in the thesis experiments were tested a minimum of 30 minutes after awakening.

Still Tassi and Muzet (2000) has pointed out that sleep inertia may occur over a longer time-course. An experimental study by Achermann and Werth (1995) highlighted that the time-course of sleep inertia follows an exponential function where reduction in performance can last for close to 1.5 hours. Jewett et al. (1999) also pointed out that based on participants' behavioural performance and subjective rating of alertness, sleep inertia can even last for 2 to 4 hours after initial awakening. Therefore it is possible that participants in the current thesis may have suffered from sleep inertia. For example, one of the limitations in Experiment 3 was the possibility that participants in the nap group may have experienced sleep inertia. This is because unlike Experiments 1 and 2, participants in Experiment 3 had faster RTs for the baseline English word-pairs in the wake condition compared to the nap condition. The above findings could indeed be due to sleep inertia effects. However, a further replication with a pre-sleep testing baseline that occurs after training would be needed to validate if results were driven by sleep inertia (i.e. difference in pre-sleep and post-sleep performance) or other experiment-specific factors.

All participants in the current thesis experiments completed the Stanford Sleepiness Scale (SSS; Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973) prior to the training (learning) and subsequent testing session. The SSS enabled the experimenter to record how alert participants felt during training and testing. Participants in Experiment 3 and 4 also completed a 2AFC RT alertness task (van der Helm et al., 2011) in addition to the SSS. There were no reported differences between training and test session or between experimental groups for all the thesis experiments, hence it is unlikely that majority of the participants had suffered from sleep inertia.

6.6.2 Circadian confounds

A weakness commonly related to sleep studies is the issue of circadian confounds when participants are tested at different time points. For example, even though all participants in Experiment 1 had an approximately 10-hour interval between training and testing, the timings of training and testing differed between the experimental groups. The wake group were trained in the morning and tested in evening whereas the sleep group were trained in the evening and tested in the morning (following day).

It is possible that the differences between the sleep and wake groups may have been confounded by factors based on time of testing. Still, apart from the newly learnt items, participants were also tested on baseline measures such as the English word-pairs and also the standardised intelligence test. The fact that sleep-associated effects only emerged for the newly learnt words highlights that the differences between the sleep and wake groups reflect the role of sleep on memory consolidation of new words. Experiment 2 adopted a nap design and hence eliminated the issue of circadian confounds between sessions as all the nap and wake training and test sessions were conducted at the same timings. Experiment 2 also found sleep-associated benefits with the participants who were trained and tested at the same time, suggesting that the findings in Experiment 1 are not likely to be confounded by time of day effects.

Still it can be argued that even though performance for well consolidated items may not differ based on time of day, the retrieval of newly learnt items may involve different mechanisms that are affected by the time of testing. Therefore, for future studies that adopt a similar design as Experiment 1, it may be theoretically important to include a group that is trained at the same time as the wake group and tested 24 hours later but at the same time as the wake group (approximately 36 hours interval between training and test).

6.7 Conclusions and future work

6.7.1 Main contributions of this thesis

This thesis makes a number of contributions to existing research on the effects of sleep on declarative learning and knowledge integration. The main contributions of the can be summarised in the three points stated below.

1. Sleep benefits memory consolidation and knowledge integration of new second language words. It is proposed that slow-wave sleep mainly helps strengthen the explicit form-meaning mappings for the new words, whereas sleep spindle activity is associated with integrating form-meaning mappings with existing knowledge.

2. Semantic distance and size congruity automaticity effects in processing new declarative memory require different time-courses and levels of sleep dependence. Due to stronger S-R mappings, semantic access of new words requires a longer time-course and greater sleep dependence for the size congruity compared to the semantic distance effect.

3. Selective factors such as strength of encoding, item saliency and nature of stimuli moderate sleep-dependent consolidation. The thesis findings suggest that sleep preferentially consolidates new words that are moderately encoded (encoding strength) over words that are very weakly or very strongly encoded. In addition, sleep-associated enhancements in automaticity increases with greater item saliency and also for second language stimuli that have more flexible first language representations (i.e. greater individual differences in existing representations).

6.7.2 Future work

Considerations for future work have been highlighted throughout this thesis, especially in this chapter. Although all the suggested avenues for future work are important and interesting, two particular areas seem especially crucial and are expanded in the two paragraphs below.

The main findings from the thesis regarding the role of slow-wave sleep and sleep spindle activity in knowledge integration of new second language words suggest that the complementary learning systems model (McClelland, McNaughton, & O'Reilly, 1995; McClelland, 2013) of word learning (Davis & Gaskell, 2009) is a useful model for understanding the development of automatic semantic access for new second language words. Further replications investigating the changes in functional connectivity before and after overnight consolidation, would be useful to examine if sleep would also be directly associated with functional changes in the brain during second language learning. It is hypothesised that slow-wave sleep and sleep spindle activity would be associated with greater functional connectivity between hippocampal and neocortical regions. By supplementing the current polysomnography and behavioural evidence with functional evidence, it would be possible to attribute a greater and more causal role of sleep of memory consolidation and knowledge integration of new word meanings.

Even though Figure 28 and 29 provide useful information about the time-course and level of sleep dependence necessary for the emergence of automatic semantic access of new word meanings, further experimental research should be conducted to look at the development of automatic semantic access over a longer time course (i.e. multiple consolidation nights). It has been predicted that greater degrees of autonomous automaticity would require greater amounts of offline consolidation. Experiment 4 has indicated that three consolidation nights was not sufficient for the emergence of autonomous automatic semantic access (physical comparisons) of new second language words. It would be useful to expand on Experiment 4 to investigate the amount of offline consolidation needed for above autonomous automaticity effects to occur. Moreover, although it is effortful to incorporate sleep polysomnography on experimental studies involving multiple consolidation nights, such data would be key in giving insights on how the role of individual sleep parameters such as slow-wave sleep and sleep spindles would change beyond a single offline consolidation opportunity.

Appendices

Appendix 1: Accuracy Measures for All Experiments

Appendix 1.1. Experiment 1

Table A1. Mean error rates (untrimmed) for Experiment 1. Standard errors are presented in parentheses.

	English		Malay	
	Wake	Sleep	Wake	Sleep
Congruity				
Congruent	.03 (.01)	.03 (.01)	.08 (.02)	.04 (.02)
Incongruent	.04 (.01)	.05 (.01)	.10 (.02)	.05 (.03)
Semantic Distance				
Small	.05 (.01)	.06 (.01)	.12 (.02)	.07 (.02)
Large	.02 (.01)	.02 (.01)	.06 (.01)	.02 (.02)

For the English comparisons, the wake and sleep group had a 4% and 4% rate of error respectively. For the Malay comparisons, the mean error rate was 9% in the wake group and 4% in the sleep group. A further breakdown of the mean error rates can be found in Table A1. The accuracy data was then arcsine square-root corrected and an ANOVA was carried out on the corrected accuracy data using the factors as the RT analysis. Similar to the RT analysis, there was a significant main effect of language where participants made more errors for the Malay compared to the English comparison task [$F(1, 27) = 7.11$, $p = .011$]. There was also an interaction between language and group [$F(1, 27) = 4.23$, $p = .049$], where the wake group made significantly more errors for the Malay compared to the English stimuli [$F(1, 13) = 10.19$, $p = .007$]. In contrast, there was no significant difference in error rates between Malay and English stimuli for the sleep group [$F(1, 13) = .21$, $p = .65$].

When analysing the Malay and English tasks separately, the findings mirrored those of the RTs to a large extent. Participants exhibited a significantly higher error rate for trials with smaller compared to larger semantic distance for both the English [$F(1,27) = 68.85$, $p < .001$] and Malay [$F(1,27) = 47.89$, $p < .001$] comparisons. Furthermore, participants made more errors for incongruent compared to congruent trials for both the English [$F(1,27) = 11.83$, $p = .002$] and Malay [$F(1,27) = 7.87$, $p = .009$] comparisons. As the main effects of congruity were only marginally significant in the RT analysis, the pattern of results for error rates highlights that the size congruity effects may be more pronounced when analysed in terms of accuracy.

All other main effects and interactions were not significant for either the Malay or English comparison tasks except the interaction between congruity, physical difference, and semantic distance [$F(1,27) = 4.73$, $p = .039$] for the English comparisons, whereby participants made significantly more errors for incongruent items with large semantic distances and smaller (1mm) physical font height differences. This interaction may have been due to the difficulty in processing such items as they were 'most different' (contrasted the most) in relation to participants' existing knowledge of animals.

Appendix 1.2. Experiment 2

Table A2. Mean error rates (untrimmed) for Experiment 2. Standard errors are presented in parentheses.

	English		Malay	
	Wake	**Sleep**	**Wake**	**Sleep**
Block 1				
Congruity				
Congruent	.04 (.01)	.05 (.01)	.06 (.01)	.05 (.01)
Incongruent	.05 (.01)	.05 (.01)	.08 (.01)	.07 (.01)
Semantic Distance				
Small	.08 (.01)	.087 (.01)	.10 (.02)	.10 (.02)
Large	.02 (.01)	.02 (.01)	.03 (.01)	.02 (.01)
Block 2				
Congruity				
Congruent	.05 (.01)	.05 (.01)	.05 (.01)	.05 (.01)
Incongruent	.06 (.01)	.06 (.01)	.06 (.01)	.06 (.01)
Semantic Distance				
Small	.09 (.01)	.09 (.01)	.09 (.02)	.09 (.02)
Large	.02 (.01)	.02 (.01)	.02 (.01)	.01 (.01)

Error rates for the English and Malay comparison tasks were arcsine square-root corrected (for untrimmed error rates see Table A2) and then analysed using an ANCOVA with the same variables as the RT analysis. Similar to the RT analysis, there was a main effect of semantic distance [$F(1, 18) = 197.66$, $p < .001$] that was present in both the English [$F(1, 18) = 190.46$, $p < .001$] and Malay tasks [$F(1, 18) = 120.46$, $p < .001$] where participants made significantly more errors for trials with small compared to large distances. In the contrast to the RT analysis, there was a main effect of congruity [$F(1, 18) = 10.83$, $p = .004$] that was driven by the Malay comparisons [$F(1, 18) = 14.56$, $p = .001$], suggesting that the above automaticity effect was reflected in terms of accuracy but not RTs. There was no main effect of congruity for the English comparisons [$F(1, 18) = 2.54$, $p = .13$]. Still it should be noted that there was no higher order interaction between language and congruity [$F(1, 18) = .73$, $p = .40$], highlighting that the congruity effect in the

Malay comparisons was not significantly greater than the effect in the English comparisons.

In addition, there was a significant three-way interaction between language, condition and training performance [$F(1, 18) = 13.53$, $p = .002$], that remained significant when analysing the English [$F(1, 18) = 4.79$, $p = .042$] and Malay trials [$F(1, 18) = 9.42$, $p = .007$] separately. Further analysis revealed three-way interaction was driven by differences in the language and training performance interaction between the nap and wake condition. There were no significant differences between training performance for the English and Malay items when the participants experienced the nap condition [$F(1, 18) = .01$, $p = .92$]. However, there was a significant interaction between language and training performance when participants experienced the wake condition [$F(1, 18) = 16.79$, $p < .001$]. Individual correlations for the wake condition revealed that, there was a significant correlation between training performance and error rates for the Malay word-pairs, $r = .55$, $p = .009$, but only a marginal significant correlation for the English word-pairs, $r = -.39$, $p = .079$. A Steiger's z-transformation affirmed that the correlation between the Malay words was significantly larger than that for the existing English words, $z = 3.09$, $p = .002$. This demonstrates that participants who did not learn the words well during training also made more errors when comparing newly learnt words if they remained awake.

Appendix 1.3. Experiment 3

Table A3. Mean error rates (untrimmed) for Experiment 3. Standard errors are presented in parentheses.

	English		Malay	
	Wake	**Nap**	**Wake**	**Nap**
Block 1				
Congruity				
Congruent	.10 (.01)	.08 (.01)	.16 (.02)	.16 (.02)
Incongruent	.11 (.02)	.09 (.01)	.18 (.02)	.20 (.02)
Semantic Distance				
Small	.15 (.02)	.12 (.01)	.23 (.02)	.25(.02)
Large	.05 (.01)	.04 (.01)	.11 (.02)	.12 (.02)
Target				
Strong			.15 (.02)	.14 (.02)
Weak			.22 (.03)	.20 (.03)
Foil				
Strong			.21 (.01)	.18 (.02)
Weak			.17 (.03)	.16 (.02)
Block 2				
Congruity				
Congruent	.11 (.01)	.10 (.02)	.13 (.02)	.13 (.02)
Incongruent	.11 (.01)	.11 (.01)	.14 (.02)	.14 (.02)
Semantic Distance				
Small	.15 (.02)	.19 (.02)	.20 (.02)	.18 (.02)
Large	.06 (.01)	.04 (.01)	.07 (.01)	.08 (.02)
Target				
Strong			.11 (.02)	.08 (.02)
Weak			.16 (.03)	.18 (.03)
Foil				
Strong			.14 (.02)	.13 (.02)
Weak			.13 (.02)	.13 (.03)

Block 3					
Congruity					
Congruent	.11 (.01)	.11 (.01)		.13 (.02)	.12 (.02)
Incongruent	.11 (.01)	.11 (.01)		.14 (.02)	.15 (.02)
Semantic Distance					
Small	.14 (.02)	.16 (.01)		.19 (.02)	.20 (.02)
Large	.07 (.01)	.05 (.01)		.07 (.01)	.06 (.02)
Target					
Strong				.11 (.02)	.11 (.01)
Weak				.16 (.03)	.15 (.02)
Foil					
Strong				.13 (.02)	.15 (.02)
Weak				.13 (.02)	.11 (.02)
Block 4					
Congruity					
Congruent	.13 (.02)	.10 (.02)		.15 (.02)	.13 (.02)
Incongruent	.12 (.01)	.11 (.01)		.17 (.02)	.15 (.02)
Semantic Distance					
Small	.18 (.02)	.16 (.02)		.23 (.02)	.21 (.02)
Large	.07 (.01)	.04 (.01)		.08 (.02)	.06 (.02)
Target					
Strong				.14 (.02)	.11 (.01)
Weak				.18 (.03)	.17 (.03)
Foil					
Strong				.17 (.02)	.14 (.02)
Weak				.15 (.02)	.13 (.02)

Note: Strength of Target and Foil was not manipulated for the English comparisons

Error rates for the English and Malay comparison tasks were arcsine square-root corrected (untrimmed error rates presented in Table 13) and then analysed using ANOVAs with the same variables as the RT analysis. Interactions with block were not presented. Participants made significantly more errors for trials with small compared to large distances in both the English [$F(1, 17) = 201.56$, < .001] and Malay [$F(1, 13) = 119.58$, $p < .001$] comparisons. In contrast, there was no difference in error rates for congruent and incongruent trials for the English [$F(1, 17) = .42$, $p = .52$] and Malay [$F(1, 13) = .52$, $p = .48$] comparisons. For the Malay comparisons, participants made significantly more errors for strong compared to weak foils [$F(1, 13) = 28.15$, $p < .001$], whereas the was no significant effect of target strength on participants' error rates [$F(1, 13) = .03$, $p = .87$]. Analysis of error rates did not reveal further significant main effects and interactions.

Appendix 1.4. Experiment 4

Table A4.1. Mean error rate (untrimmed) for the semantic comparisons in Experiment 4. Standard errors are presented in parentheses.

	English				Malay			
	Day 1	Day 2	Day 3	Day 4	Day 1	Day 2	Day 3	Day 4
Block 1								
Congruity								
Congruent	.05 (.01)	.07 (.01)	.06 (.01)	.06 (.01)	.12 (.02)	.09 (.04)	.06 (.02)	.06 (.01)
Incongruent	.06 (.01)	.09 (.01)	.08 (.01)	.07 (.01)	.09 (.02)	.10 (.02)	.07 (.01)	.07 (.01)
Semantic Distance								
Small	.09 (.01)	.12 (.01)	.11 (.01)	.11 (.01)	.15 (.02)	.14 (.02)	.11 (.02)	.11 (.02)
Large	.03 (.01)	.04 (.01)	.04 (.01)	.03 (.01)	.06 (.01)	.05 (.01)	.02 (.01)	.02 (.01)
Block 2								
Congruity								
Congruent	.06 (.01)	.05 (.01)	.07 (.01)	.07 (.01)	.12 (.03)	.13 (.03)	.07 (.02)	.07 (.02)
Incongruent	.10 (.01)	.12 (.02)	.13 (.02)	.11 (.02)	.13 (.02)	.12 (.02)	.09 (.01)	.10 (.02)
Semantic Distance								
Small	.13 (.02)	.14 (.01)	.16 (.01)	.15 (.02)	.20 (.03)	.17 (.03)	.13 (.02)	.13 (.02)

Large	.03	.03	.04	.03	.07	.08	.03	.03
	(.01)	(.01)	(.01)	(.01)	(.02)	(.03)	(.01)	(.01)

Accuracy rates (see Table A4.1 for untrimmed error rates) for the semantic comparisons were arcsine square-root corrected and subjected to an ANOVA with identical factors as the RT analysis. Interactions with presentation block were not reported except for cases of theoretical interest. Unlike the RT analysis, there was no significant main effect of language [$F(1, 21) = .44$, $p = .51$] or day [$F(3, 63) = 2.66$, $p = .73$]. Still, there was a significant interaction between language and day [$F(3, 63) = 6.73$, $p < .001$], which was driven by a significant main effect of test day in the Malay trials [$F(3, 57) = 3.25$, $p = .028$], but not the English trials [$F(3, 57) = 1.45$, $p = .24$]. This suggests that learning related improvements in accuracy occurred for the Malay trials where participants made less errors with additional time and consolidation (nights), whereas there were no equivalent relationships for the English trials. Similar to the RT analysis, pair-wise comparisons were carried out between the consecutive days, to gain insights on the incremental benefits of additional consolidation nights (p-values were bonferroni corrected for the number of comparisons made). This resulted in a threshold p-value of .0167 after Bonferroni corrections. Pair-wise comparisons revealed that participants made significantly less errors on day 3 as compared to day 2 [$F(1, 21) = 7.59$, $p = .012$]. However, the difference in accuracy rates between day 1 and day 2, and the difference between day 3 and day 4 were not significant.

There was a significant main effect of congruity [$F(1, 21) = 6.67$, $p = .017$], where for both the English [$F(1, 19) = 8.94$, $p = .007$] and Malay [$F(1, 19) = 7.16$, $p = .015$] trials, participants made more errors for the incongruent compared to congruent word-pairs. The main effect of semantic distance was also significant [$F(1, 21) = 333.69$, $p < .001$] where for both the English [$F(1, 19) = 160.83$, $p < .001$] and Malay [$F(1, 19) = 190.20$, $p < .001$] trials, participants made less errors for word-pairs with large compared to small distances.

There was a significant main effect of block [$F(1, 21) = 17.04$, $p < .001$] where for both the English [$F(1, 19) = 9.55$, $p = .006$] and Malay [$F(1, 19) = 5.95$, $p = .025$] word-pairs, participants made significantly more errors in the second compared to first block of trials. However, unlike the RT analysis, there was no significant interaction between language and block [$F(1, 21) = .001$, $p = .98$]. There were no other significant interactions for the within-participant variables except the interaction between language, congruity and distance [$F(1, 21) = 5.68$, $p = .027$]. Upon further analysis, the three-way interaction was driven by the point that for word-pairs with small distances, participants made significantly more errors for the incongruent compared to congruent English but not Malay trials [$F(1, 21) = 7.01$, $p = .015$]. There were no equivalent differences for word-pairs with large distances or when considering other 2-way interactions (all p's >.38). There was also no main effect of the between-participant task order variable [$F(1, 19) = .096$, $p = .76$].

Table A4.2. Mean error rate (untrimmed) for the physical comparisons in Experiment 4. Standard errors are presented in parentheses.

	English				Malay			
	Day 1	Day 2	Day 3	Day 4	Day 1	Day 2	Day 3	Day 4
Block 1								
Congruity								
Congruent	.01 (.01)	.02 (.01)	.03 (.01)	.04 (.01)	.03 (.01)	.04 (.01)	.04 (.01)	.04 (.01)
Incongruent	.04 (.01)	.04 (.01)	.05 (.01)	.06 (.01)	.03 (.01)	.03 (.01)	.03 (.01)	.04 (.01)
Semantic Distance								
Small	.03 (.01)	.03 (.01)	.04 (.01)	.05 (.01)	.03 (.01)	.03 (.01)	.03 (.01)	.05 (.01)
Large	.03 (.01)	.04 (.01)	.04 (.01)	.05 (.01)	.03 (.01)	.03 (.01)	.03 (.01)	.03 (.01)

Block 2								
Congruity								
Congruent	.04	.03	.04	.04	.02	.02	.04	.04
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
Incongruent	.03	.04	.03	.04	.02	.03	.03	.04
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
Semantic Distance								
Small	.04	.03	.03	.05	.02	.03	.04	.05
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
Large	.03	.04	.03	.03	.02	.02	.03	.04
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)

Mean accuracy rates (see Table A4.2 for untrimmed error rates) for the physical comparisons were arcsine square-root corrected and subjected to an ANOVA with identical factors as the RT analysis. Interactions with presentation block were not reported except for cases of theoretical interest. There was a marginally significant effect of language [$F(1, 21) = 3.77$, $p = .066$] for the combined ANOVA, were participants made more error for the Malay compared to English word-pairs. There was a main effect of day [$F(3, 63) = 3.32$, $p = .025$], where participants made more errors with each test day. However, when considering the languages separately, neither the English [$F(3, 57) = 2.09$, $p = .11$] nor Malay [$F(3, 57) = 1.92$, $p = .14$] word-pairs displayed a main effect of day.

There was marginal main effect of congruity [$F(1, 21) = 3.77$, $p = .066$] and also an interaction between language and congruity [$F(1, 21) = 5.54$, $p = .028$], where the main effect of congruity in the English trials [$F(1, 19) = 7.85$, $p = .011$] as participants made significantly more errors for incongruent compared to congruent word-pairs. In contrast, participants error rates in the Malay trials were not significantly affected by congruity [$F(1, 19) = .19$, $p = .89$]. There was an interaction between congruity and distance [$F(1, 21) = 6.73$, $p = .017$], where participants made significantly less errors for congruent word-pairs with large semantic distances that for other word-pairs.

There was no significant main effect of semantic distance [$F(1, 21) = 2.83$, $p > .05$] or block [$F(1, 21) = .54$, $p = .11$]. There were no other significant interactions between the within-participant variables and no between-participant of task order [$F(1, 21) = 1.87$, $p = .19$].

Appendix 1.5. Experiment 5

Table A5. Mean error rate (untrimmed) for Experiment 5. Standard errors are presented in parentheses.

	Existing Numerals				New Numerals			
	Nap	Wake	Overnight	Immediate	Nap	Wake	Overnight	Immediate
Congruity								
Congruent	.02	.01	.01	.02	.04	.06	.06	.06
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
Neutral	.02	.02	.03	.02	.05	.08	.08	.05
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
Incongruent	.05	.08	.08	.08	.05	.09	.09	.09
	(.01)	(.01)	(.01)	(.01)	(.02)	(.02)	(.02)	(.01)
Semantic Distance								
1	.05	.06	.07	.06	.07	.13	.14	.12
	(.01)	(.01)	(.01)	(.01)	(.02)	(.02)	(.02)	(.02)
2	.03	.04	.03	.04	.05	.07	.05	.06
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
4	.01	.02	.01	.02	.02	.03	.02	.02
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)

Accuracy rates (see Table A5 for untransformed error rates) were arcsine square-root transformed and analysed using ANOVAs with the same variables as the RT analysis. Similar to the RT analysis, there was a significant main effect of numeral type [$F(1, 55) = 116.47$, $p < .001$] where participants responded with significantly less errors for the existing numeral-pairs as compared to the new numeral-pairs. There was no main effect of experimental group [$F(3, 55) = .88$, $p = .46$] or presentation block [$F(1, 55) = .15$, $p = .70$].

Results also revealed a main effect of congruity [$F(2, 110) = 56.55$, $p < .001$]. For both the existing [$F(2, 100) = 80.96$, $p < .001$] and new numerals [$F(2, 110) = 10.80$, $p < .001$], participants made most errors for the incongruent trials, followed by the neutral trials and least errors for congruent trials. Similar to the RT analysis, there was also a significant interaction between numeral type and congruity [$F(1, 110) = 28.07$, $p < .001$]. Further analysis revealed that participants made less errors for the existing as compared to new numeral-pairs for congruent [$F(1, 55) = 90.73$, $p < .001$] and neutral [$F(1, 55) = 125.06$, $p < .001$] trials. In contrast, there was no different in accuracy performance between the existing and new numeral-pairs for incongruent trials [$F(1, 55) = 1.05$, $p = .31$].

There was a main effect of distance [$F(2, 110) = 193.17$, $p < .001$], whereby participants made more errors for trials with smaller compared to larger numerical distances for the existing [$F(2, 110) = 68.10$, $p < .001$], and new [$F(2, 110) = 164.47$, $p < .001$] numeral pairs. There was also a significant interaction between numeral type, congruity and distance [$F(4, 220) = 6.68$, $p < .001$]. There three-way interaction was driven by a two-way interaction between type and congruity that was significant between numeral-pairs that differed by 1 distance [$F(2, 110) = 20.42$, $p < .001$] or by a distance of 2 [$F(2, 110) = 16.02$, $p < .001$], but not significant in pairs that differed by a distance of 4 [$F(2, 110) = .32$, $p = .73$]. For the numeral-pairs that differed by a distance of 1 or 2, participants made significantly more errors for existing compared to new numeral-pairs in the congruent and neutral trials, but not the incongruent trials.

There were no further statistically significant interactions except the interaction between distance and condition [$F(2, 110) = 2.55$, $p = .024$], congruity and condition [$F(2, 110) = 2.44$, $p = .030$], numeral type and distance [$F(2, 110) = 23.00$, $p < .001$], the interaction between congruity and distance [$F(4, 220) = 3.55$, $p = .008$] and also the interaction between numeral type, congruity, distance and condition [$F(12, 220) = 1.99$, $p = .026$]. Further analyses of all the above interactions did not reveal any theoretically significant differences.

Appendix 2. Experiment 1: List of experimental stimuli

English animal names and corresponding Malay translations. (Mean size norm rating and standard deviations presented in parentheses)

	SET A		SET B	
	English	**Malay**	**English**	**Malay**
Small (M = 1.41, SD = 0.45)	ANT	SEMUT	FLEA	KUTU
	BEE	LEBAH	SNAIL	SIPUT
	FROG	KATAK	CRAB	KETAM
Medium (M = 3.81, SD = 0.66)	DUCK	ITIK	FOX	RUBAH
	EAGLE	HELANG	SHEEP	BIRI
	CAT	KUCING	DOG	ANJING
Large (M = 6.63, SD = 0.55)	COW	LEMBU	MULE	BAGAL
	HORSE	KUDA	CAMEL	UNTA
	MOOSE	RUSA	LION	SINGA

Appendix 3. Experiment 1: Summary of means for the congruity, semantic distance and physical font difference factors.

	English		Malay	
	Wake	**Sleep**	**Wake**	**Sleep**
Congruity (Co)				
Congruent	904 (45)	918 (50)	1338 (82)	1157 (72)
Incongruent	912 (49)	946 (45)	1360 (82)	1187 (73)
Semantic Distance (Dist)				
Small	953 (47)	983 (48)	1403 (85)	1237 (76)
Large	863 (47)	881 (47)	1294 (79)	1107 (69)
Physical Font Difference (Phy)				
Small (1mm)	914 (47)	931 (47)	1357 (84)	1166 (73)
Large (4mm)	903 (47)	933 (47)	1341 (80)	1178 (72)
Co x Dist				
Congruent				
Small Dist	950 (46)	966 (52)	1383 (85)	1231 (76)
Large Dist	859 (46)	871 (48)	1293 (80)	1082 (68)
Incongruent				
Small Dist	957 (50)	1000 (44)	1423 (85)	1243 (76)
Large Dist	868 (49)	892 (48)	1296 (80)	1131 (72)
Co x Phy				
Congruent				
Small Phy	909 (45)	911 (48)	1345 (81)	1156 (71)
Large Phy	900 (46)	925 (52)	1331 (84)	1157 (72)
Incongruent				
Small Phy	919 (50)	952 (47)	1369 (88)	1176 (75)
Large Phy	906 (49)	940 (44)	1351 (77)	1198 (72)

Dist x Phy					
Small Dist					
Small Phy	951 (46)	983 (48)		1413 (89)	1237 (80)
Large Phy	877 (49)	880 (48)		1300 (80)	1095 (66)
Large Dist					
Small Phy	956 (48)	982 (48)		1392 (81)	1237 (72)
Large Phy	850 (46)	883 (48)		1289 (80)	1119 (72)
Co x Dist x Phy					
Congruent					
Small Dist					
Small Phy	939 (44)	956 (49)		1400 (86)	1299 (79)
Large Phy	961 (49)	975 (57)		1366 (88)	1233 (74)
Large Dist					
Small Phy	879 (49)	867 (49)		1290 (79)	1083 (65)
Large Phy	838 (45)	874 (49)		1295 (82)	1081 (72)
Incongruent					
Small Dist					
Small Phy	964 (50)	1011 (49)		1427 (97)	1246 (83)
Large Phy	950 (51)	989 (42)		1419 (76)	1240 (71)
Large Dist					
Small Phy	874 (50)	892 (48)		1311 (82)	1107 (70)
Large Phy	862 (49)	892 (49)		1282 (80)	1156 (75)

Appendix 4. Experiment 2: List of experiment stimuli

	Set 1				Set 2			
	Group A		Group B		Group A		Group B	
Size Ranking	Malay	English	Malay	English	Malay	English	Malay	English
Small	MINAT	ANT	MAKNA	FLEA	IKLAN	FLY	WAJIB	BEETLE
	EMPAT	MOTH	YURAN	BEE	JIRAN	WORM	MINAT	SHRIMP
	IKLAN	FROG	WAKIL	SNAIL	KERJA	LIZZARD	EMPAT	CRAB
Medium	INGAT	PARROT	JAWAB	DUCK	YURAN	TURTLE	INGAT	HEN
	JIRAN	CAT	BETUL	SKUNK	WAKIL	GOOSE	TIDAK	OWL
	KERJA	MONKEY	MURID	EAGLE	JAWAB	SWAN	MAKNA	OTTER
Large	NASIB	ZEBRA	WAKAF	DEER	BETUL	LION	SIBUK	DONKEY
	TIDAK	HORSE	HIDUP	MULE	WAKAF	PANDA	MURID	TIGER
	SIBUK	BEAR	WAJIB	MOOSE	NASIB	CAMEL	HIDUP	COW

Appendix 5. Experiment 3: List of experimental stimuli.

Words in the strong encoding condition are depicted in red.

	Set 1				Set 2			
	Group A		Group B		Group A		Group B	
Size Ranking	Malay	English	Malay	English	Malay	English	Malay	English
Small	MINAT	ANT	MAKNA	FLEA	IKLAN	FLY	WAJIB	BEETLE
	EMPAT	MOTH	YURAN	BEE	JIRAN	WORM	MINAT	SHRIMP
	IKLAN	FROG	WAKIL	SNAIL	KERJA	LIZZARD	EMPAT	CRAB
	WAKTU	MOUSE	KARYA	BAT	KECIL	CHICK	AKHIR	TOAD
Medium	INGAT	PARROT	JAWAB	DUCK	YURAN	TURTLE	INGAT	HEN
	JIRAN	CAT	BETUL	SKUNK	WAKIL	GOOSE	TIDAK	OWL
	KERJA	MONKEY	MURID	EAGLE	JAWAB	SWAN	MAKNA	OTTER
	CUKAI	FOX	ENCIK	DOG	EHWAL	KOALA	NILAI	LAMB
Large	NASIB	ZEBRA	WAKAF	DEER	BETUL	LION	SIBUK	DONKEY
	TIDAK	HORSE	HIDUP	MULE	WAKAF	PANDA	MURID	TIGER
	SIBUK	BEAR	WAJIB	MOOSE	NASIB	CAMEL	HIDUP	COW
	MESRA	GOAT	HUKUM	ELK	CIKGU	SHEEP	SIJIL	WOLF

Appendix 6. Experiment 4: Sleep parameters for participants across consolidation nights.

Parentheses denote the standard deviation of the means.

Sleep Parameter	Mean time in minutes	Time as a % of total sleep time
Baseline Night		
Total sleep time	423 (55)	
Wake time after sleep onset	10 (14)	
Light Sleep	212 (50)	50 (7)
Deep Sleep	91 (24)	22 (7)
REM Sleep	121 (32)	28 (6)
Night 1		
Total sleep time	447 (76)	
Wake time after sleep onset	7 (10)	
Light Sleep	231 (39)	52 (8)
Deep Sleep	82 (22)	18 (4)
REM Sleep	135 (52)	30 (8)
Night 2		
Total sleep time	425 (67)	
Wake time after sleep onset	16 (29)	
Light Sleep	213 (52)	50 (9)
Deep Sleep	89 (22)	21 (6)
REM Sleep	124 (39)	29 (7)
Night 3		
Total sleep time	409 (58)	
Wake time after sleep onset	7 (9)	
Light Sleep	200 (57)	49 (11)
Deep Sleep	98 (29)	24 (7)
REM Sleep	111 (37)	27 (8)

References

Achermann, P., Werth, E., Dijk, D. J., & Borbely, A. A. (1995). Time course of sleep inertia after nighttime and daytime sleep episodes. *Archives italiennes de biologie, 134(1),* 109-119.

Aitchison, J. (2003). *Words in the mind: an introduction to the mental lexicon*. Wiley-Blackwell.

Alger, S. E., Lau, H., & Fishbein, W. (2012). Slow wave sleep during a daytime nap is necessary for protection from subsequent interference and long-term retention. *Neurobiology of Learning and Memory, 98,* 188–196.

Algina, J. (1982). Remarks on the analysis of covariance in repeated measures designs. *Multivariate Behavioral Research, 17,* 117–130.

Alison G, & Meltzoff, A. (1987). The Development of Categorization in the Second Year and Its Relation to Other Cognitive and Linguistic Developments. *Child Development, 58*, 1523–1531.

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature reviews. Neuroscience, 9,* 278–291.

Ansari, D., Garcia, N., Lucas, E., Hamon, K., & Dhital, B. (2005). Neural correlates of symbolic number processing in children and adults. *Neuroreport, 16*, 1769–1773.

Backhaus, J., & Junghanns, K. (2006). Daytime naps improve procedural motor memory. *Sleep Medicine, 7,* 508–512.

Balota, D. a, Yap, M. J., Cortese, M. J., Hutchison, K. a, Kessler, B., Loftis, B., … Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445–459.

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development, 79(4),* 1016-1031.

Borbély, A. A., & Achermann, P. (1999). Sleep homeostasis and models of sleep regulation. *Journal of Biological Rhythms, 14,* 557–568.

Borbély, A. A., Achermann, P., Trachsel, L., & Tobler, I. (1989). Sleep Initiation and Initial Sleep Intensity: Interactions of Homeostatic and Circadian Mechanisms. *Journal of Biological Rhythms, 4,* 37 –48.

Bromiley, P. A., & Thacker, N. A. (2002). The effects of an arcsin square root transform on a binomial distributed quantity. *Tina memo, 7*, 1-8.

Brown, G. D. a, Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114,* 539–576.

Brown, G., & Lewandowsky, S. (2010). Forgetting in memory models: Arguments against trace decay and consolidation failure. *Current Issues in Memory,* 1-40.

Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. Philosophical transactions of the Royal Society of London. *Biological Sciences, 364,* 1831–1840.

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

Buzsáki, G. (1989). Two-stage model of memory trace formation: A role for "noisy" brain states. *Neuroscience, 31*, 551–570.

Buzsáki, G. (1996). The hippocampo-neocortical dialogue. *Cerebral Cortex, 6*, 81–92.

Canli, T., Sivers, H., Thomason, M. E., Whitfield-Gabrieli, S., Gabrieli, J. D. E., & Gotlib, I. H. (2004). Brain activation to emotional words in depressed vs healthy subjects. *Neuroreport, 15*, 2585258–8.

Carr, T. H. (1992). Automaticity and Cognitive Anatomy: Is Word Recognition "Automatic"? *The American Journal of Psychology, 105*, 201-237.

Carskadon, M. A., & Dement, W. C. (2000). Normal human sleep: an overview. *Principles and practice of sleep medicine, 4*, 13-23.

Cirelli, C., & Tononi, G. (2008). Is Sleep Essential? *PLoS Biol, 6,* e216.

Cohen Kadosh, R., & Henik, A. (2006). A Common Representation for Semantic and Physical Properties. *Experimental Psychology (formerly "Zeitschrift für Experimentelle Psychologie"), 53,* 87–94.

Cohen Kadosh, R., Henik, A., Rubinsten, O., Mohr, H., Dori, H., van de Ven, V., … Linden, D. E. J. (2005). Are numbers special? The comparison systems of the human brain investigated by fMRI. *Neuropsychologia, 43,* 1238–1248.

Cohen Kadosh, R., Lammertyn, J., & Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Progress in Neurobiology, 84,* 132–147.

Cohen Kadosh, R., Soskic, S., Iuculano, T., Kanai, R., & Walsh, V. (2010). Modulating Neuronal Activity Produces Specific and Long-Lasting Changes in Numerical Competence. *Current Biology, 20,* 2016–2020.

Corkin, S. (2002). What's new with the amnesic patient H.M.? Nature reviews. *Neuroscience, 3*, 153–160.

Cox & Calderon (2014). Speedup versus Automatization. In Connor-Linton, J., & Amoroso, L. W. (Eds.). *Measured Language: Quantitative Studies of Acquisition, Assessment, and Variation* (pp. 111- 124). Georgetown University Press.

Crutch, S., & Warrington, E. (2002). Preserved calculation skills in a case of semantic dementia. *Cortex*, 389–399.

Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience, 21*, 803–820.

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *364,* 3773–3800.

Dehaene, S. (1996). The Organization of Brain Activations in Number Comparison: Event-Related Potentials and the Additive-Factors Method. *Journal of Cognitive Neuroscience, 8,* 47–68.

Delaney, H., & Maxwell, S. (1981). On using analysis of covariance in repeated measures designs. *Multivariate Behavioral Research, 16,* 101–123.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134,* 9–21.

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

Diekelmann, S., & Born, J. (2010b). Slow-wave sleep takes the leading role in memory reorganization. *Nature Reviews Neuroscience, 11,* 218.

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

Dumay, N., & Gaskell, M. G. (2007). Sleep-Associated Changes in the Mental Representation of Spoken Words. *Psychological Science,* 18, 35 –39.

Eimer, M. (1995). Stimulus-response compatibility and automatic response activation: evidence from psychophysiological studies. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 837-854.

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,* 7723 –7728.

Ellenbogen, J. M., Hulbert, J. C., Stickgold, R., Dinges, D. F., & Thompson-Schill, S. L. (2006). Interfering with Theories of Sleep and Memory: Sleep, Declarative Memory, and Associative Interference. *Current Biology, 16,* 1290–1294.

Ellenbogen, J. M., Payne, J. D., & Stickgold, R. (2006). The role of sleep in declarative memory consolidation: passive, permissive, active or none? *Current Opinion in Neurobiology, 16,* 716–722.

Ferrarelli, F., Huber, R., Peterson, M. J., Massimini, M., Murphy, M., Riedner, B. a, … Tononi, G. (2007). Reduced sleep spindle activity in schizophrenia patients. *The American Journal of Psychiatry, 164,* 483–492.

Flehmig, H. C., Steinborn, M., Langner, R., Scholz, A., & Westhoff, K. (2007). Assessing intraindividual variability in sustained attention: Reliability, relation to speed and accuracy, and practice effects. *Psychology Science, 49*(2), 132-149.

Fogel, S. M., & Smith, C. T. (2006). Learning-dependent changes in sleep spindles and Stage 2 sleep. *Journal of Sleep Research, 15*, 250–255.

Fogel, S. M., & Smith, C. T. (2011). The function of the sleep spindle: a physiological index of intelligence and a mechanism for sleep-dependent memory consolidation. *Neuroscience and Biobehavioral Reviews, 35,* 1154–1165.

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,* 48–61.

Foltz, G. S., Poltrock, S. E., & Potts, G. R. (1984). Mental comparison of size and magnitude: size congruity effects. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 10,* 442–453.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers,35*(1), 116-124.

Francis, W. S. (1999). Cognitive integration of language and memory in bilinguals: semantic representation. *Psychological Bulletin, 125,* 193–222.

Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nature Reviews Neuroscience, 6,* 119–130.

Fukkink, R. G., Hulstijn, J., & Simis, A. (2005). Does Training in Second-Language Word Recognition Skills Affect Reading Comprehension? An Experimental Study. *The Modern Language Journal*, *89*(1), 54-75.

Gais, S., & Born, J. (2004). Low acetylcholine during slow-wave sleep is critical for declarative memory consolidation. *Proceedings of the National Academy of Sciences of the United States of America, 101,* 2140 –2144.

Gais, S., Mölle, M., Helms, K., & Born, J. (2002). Learning-dependent increases in sleep spindle density. *The Journal of Neuroscience,* 22, 6830–6634.

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

Ganger, J., & Brent, M. R. (2004). Reexamining the Vocabulary Spurt. *Developmental Psychology, 40,* 621–632.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89,* 105–132.

Gebuis, T., Cohen Kadosh, R., de Haan, E., & Henik, A. (2009). Automatic quantity processing in 5-year olds and adults. *Cognitive Processing, 10,* 133–142.

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

Glenberg, a M., & Swanson, N. G. (1986). A temporal distinctiveness theory of recency and modality effects. Journal of experimental psychology. *Learning, Memory, and Cognition, 12,* 3–15.

Gollan, T. H., & Ferreira, V. S. (2009). Should I stay or should I switch? A cost–benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 640-665.

Griessenberger, H., Heib, D. P. J., Kunz, A. B., Hoedlmoser, K., & Schabus, M. (2013). Assessment of a wireless headband for automatic sleep scoring. *Sleep and Breathing*, *17*(2), 747-752.

Gumenyuk, V., Roth, T., Korzyukov, O., Jefferson, C., Bowyer, S., & Drake, C. L. (2011). Habitual short sleep impacts frontal switch mechanism in attention to novelty. *Sleep, 34,* 1659–1670.

Henderson, L., Weighall, A., & Gaskell, G. (2013). Learning new vocabulary during childhood: effects of semantic training on lexical consolidation and integration. *Journal of Experimental Child Psychology, 116,* 572–592.

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

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

Iber C., Ancoli-Israel S., Quan S.F. (2007). *The AASM manual for the scoring of sleep and associated events: rules, terminology, and technical specifications.* The American Academy of Sleep Medicine1st ed. Westchester, Ill, American Academy of Sleep Medicine.

Jewett, M. E., Wyatt, J. K., Ritz-De Cecco, a, Khalsa, S. B., Dijk, D. J., & Czeisler, C. A. (1999). Time course of sleep inertia dissipation in human performance and alertness. *Journal of Sleep Research, 8,* 1–8.

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

Karni, A., & Giuseppe, B. (1997). Learning perceptual skills: Behavioral probes into adult cortical plasticity. *Current Opinion in Neurobiology,* 9–13.

Kaufmann, L., Koppelstaetter, F., Siedentopf, C., Haala, I., Haberlandt, E., Zimmerhackl, L.-B., … Ischebeck, A. (2006). Neural correlates of the number–size interference task in children. *Neuroreport, 17,* 587–591.

Klimesch, W., Pfurtscheller, G., Mohl, W., & Schimke, H. (1990). Event-related desynchronization, ERD-mapping and hemispheric differences for words and numbers. *International Journal of Psychophysiology*, *8*(3), 297-308.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus-response compatibility--a model and taxonomy. *Psychological review*, *97*(2), 253-270.

Kuriyama, K., Stickgold, R., & Walker, M. P. (2004). Sleep-dependent learning and motor-skill complexity. *Learning & Memory*, *11*(6), 705-713.

Lahl, O., Wispel, C., Willigens, B., & Pietrowsky, R. (2008). An ultra short episode of sleep is sufficient to promote declarative memory performance. *Journal of Sleep Research, 17,* 3–10.

Lau, H., Alger, S. E., & Fishbein, W. (2011). Relational memory: a daytime nap facilitates the abstraction of general concepts. *PloS one, 6,* e27139.

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

Le Clec'H, G., Dehaene, S., Cohen, L., Mehler, J., Dupoux, E., Poline, J. B., … Le Bihan, D. (2000). Distinct cortical areas for names of numbers and body parts independent of language and input modality. *NeuroImage, 12,* 381–391.

Lewandowsky, S., Ecker, U. K. H., Farrell, S., & Brown, G. D. a. (2012). Models of cognition and constraints from neuroscience: A case study involving consolidation. *Australian Journal of Psychology, 64,* 37–45.

Marshall, L., Helgadóttir, H., Mölle, M., & Born, J. (2006). Boosting slow oscillations during sleep potentiates memory. *Nature, 444,* 610–613.

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,* 185–188.

McClelland, J., & Goddard, N. (1996). Considerations arising from a complementary learning systems perspective on hippocampus and neocortex. *Hippocampus, 6,* 654–665.

McClelland, J. L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General*, *142*(4), 1190-1210.

McClelland, J., McNaughton, B., & O'Reilly, R. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419–457.

McGaugh, J. L. (2000). Memory--a Century of Consolidation. *Science, 287,* 248–251.

McMurray, B. (2007). Defusing the Childhood Vocabulary Explosion. *Science, 317,* 631.

Mednick, S. C., & Alaynick, W. a. (2010). Comparing Models of Sleep-dependent Memory Consolidation. *Journal of Experimental & Clinical Medicine, 2,* 156–164.

Mednick, S. C., McDevitt, E. A., Walsh, J. K., Wamsley, E., Paulus, M., Kanady, J. C., & Drummond, S. P. (2013). The critical role of sleep spindles in hippocampal-dependent memory: A pharmacology study. *The Journal of Neuroscience, 33*(10), 4494-4504.

Mednick, S., Cai, D., Kanady, J., & Drummond, S. (2008). Comparing the benefits of caffeine, naps and placebo on verbal, motor and perceptual memory. *Behavioural Brain Research, 193,* 79–86.

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

Moyer, R. S. (1973). Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception & Psychophysics, 13,* 180–184.

Myers, J. H., & Alpert, M. I. (1977). Semantic confusion in attitude reseach: Salience vs. Importance vs. Determinance. *Advances for Consumer Research, 4,* 106–110.

Nishida, M., & Walker, M. P. (2007). Daytime Naps, Motor Memory Consolidation and Regionally Specific Sleep Spindles. *PLoS one, 2,* e341.

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,* 505–510.

Paivio, A. (1975). Perceptual comparisons through the mind's eye. *Memory & Cognition, 3,* 635–647.

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., Chambers, A. M., & Kensinger, E. a. (2012). Sleep promotes lasting changes in selective memory for emotional scenes. *Frontiers in Integrative Neuroscience, 6,* 1-11.

Payne, J. D., Tucker, M. a, Ellenbogen, J. M., Wamsley, E. J., Walker, M. P., Schacter, D. L., & Stickgold, R. (2012). Memory for semantically related and unrelated declarative information: the benefit of sleep, the cost of wake. *PloS one*, *7,* e33079.

Peigneux, P., Laureys, S., Delbeuck, X., & Maquet, P. (2001). Sleeping brain, learning brain. The role of sleep for memory systems. *Neuroreport, 12,* A111–124.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific studies of reading, 11*(4), 357-383.

Perlman, A., & Tzelgov, J. (2006). Interactions between encoding and retrieval in the domain of sequence-learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(1), 118-130.

Plihal, W., & Born, J. (1997). Effects of Early and Late Nocturnal Sleep on Declarative and Procedural Memory. *Journal of Cognitive Neuroscience, 9,* 534–547.

Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis, *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 785-794.

Rasch, B., & Born, J. (2007). Maintaining memories by reactivation. Current Opinion in *Neurobiology, 17,* 698–703.

Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Reviews, 93,* 681–766.

Rasch, B., Büchel, C., Gais, S., & Born, J. (2007). Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science,* 1426–1429.

Rechtschaffen, A., & Kales, A. (1968). *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*. UCLA Brain Information Service. Brain Research Institute, Los Angeles.

Rodd, J. M., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, M. G., & Davis, M. H. (2012). Learning new meanings for old words: effects of semantic relatedness. *Memory & cognition, 40*, 1095–1108.

Rubinsten, O., & Henik, A. (2002). Is an ant larger than a lion? *Acta Psychologica,* 111, 141–154.

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.

Rudoy, J., Voss, J., Westerberg, C., & Paller, K. (2009). Strengthening individual memories by reactivating them during sleep. *Science,* 326, 1079.

Saeed, M. M., Keens, T. G., Stabile, M. W., Bolokowicz, J., & Davidson Ward, S. L. (2000). Should children with suspected obstructive sleep apnea syndrome and normal nap sleep studies have overnight sleep studies? *Chest, 118,* 360–365.

Saletin, J. M., & Walker, M. P. (2012). Nocturnal mnemonics: sleep and hippocampal memory processing. *Frontiers in Neurology, 3(*59), 1-12.

Sanders, M. S., & McCormick, E. J. (1987). *Human factors in engineering and design* . McGraw-Hill book company.

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,* 1479–1485.

Schabus, M., Hödlmoser, K., Pecherstorfer, T., & Klösch, G. (2005). Influence of midday naps on declarative memory performance and motivation. *Somnologie*, *9*(3), 148-153.

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.

Schwarz, W., & Heinze, H. J. (1998). On the interaction of numerical and size information in digit comparison: a behavioral and event-related potential study. *Neuropsychologia, 36,* 1167–1179.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. 1957. *The Journal of Neuropsychiatry and Clinical Neurosciences, 12,* 103–113.

Segalowitz, N. (1999). *Automaticity and Second Languages.* The Handbook of Second Language Acquisition, 27, 382-391.

Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled Performance, Practice, and the Differentiation of Speed-up from Automatization Effects: Evidence from Second Language Word Recognition. *Applied Psycholinguistics, 14,* 369–385.

Segalowitz, S. J., Segalowitz, N. S., & Wood, A. G. (1998). Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics, 19,* 53-67.

Sereno, S. C., O'Donnell, P. J., & Sereno, M. E. (2009). Size matters: bigger is faster. *The Quarterly journal of experimental psychology 62,* 1115–1122.

Shambroom, J. R., Fábregas, S. E., & Johnstone, J. (2012). Validation of an automated wireless system to monitor sleep in healthy adults. *Journal of Sleep Research, 21,* 221–230.

Silber, M. H., Ancoli-Israel, S., Bonnet, M. H., Chokroverty, S., Grigg-Damberger, M. M., Hirshkowitz, M., ... & Iber, C. (2007). The visual scoring of sleep in adults. *Journal of Clinical Sleep Medicine*, *3*(2), 121-131.

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of Learning and Memory, 82,* 171–177.

Squire, L. R. (2011). Declarative and Nondeclarative Memory: Multiple Brain Systems Supporting Learning and Memory. *Journal of Cognitive Neuroscience, 4,* 232–243.

Squire, L. R., Knowlton, B., & Musen, G. (1993). The Structure and Organization of Memory. *Annual Review of Psychology, 44,* 453–495.

Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences, 93,* 13515 –13522.

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature,* 437, 1272–8.

Stickgold, R. (2009). How do I remember? Let me count the ways. *Sleep Medicine Reviews*, *13*(5), 305-308.

Stickgold, R., & Walker, M. (2007). Sleep-dependent memory consolidation and reconsolidation. *Sleep Medicine, 8,* 331–343.

Stickgold, R., & Walker, M. P. (2005). Memory consolidation and reconsolidation: what is the role of sleep? *Trends in Neurosciences*, *28,* 408–415.

Stroop, J. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology,* 643–661.

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,* 756–761.

Talamini, L. M., Nieuwenhuis, I. L. C., Takashima, A., & Jensen, O. (2008). Sleep directly following learning benefits consolidation of spatial associative memory. *Learning & Memory,* 15, 233–237.

Tamminen, J., & Gaskell, M. G. (2013). Novel word integration in the mental lexicon: Evidence from unmasked and masked semantic priming. *The Quarterly Journal of Experimental Psychology*, *66*(5), 1001-1025.

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. *The Journal of Neuroscience*, *30*(43), 14356-14360.

Tassi, P., & Muzet, A. (2000). Sleep inertia. *Sleep Medicine Reviews, 4,* 341–353.

Tham, E. K., Lindsay, S., & Gaskell, M. G. (2011, January). Effects of Sleep on Knowledge Integration and Automaticity of Processing. In *SLEEP* (Vol. 34, pp. A89-A89). One Westbrook Corporate CTR, STE 920, Westchester, IL 60154 USA: American Academy of Sleep Medicine.

Thioux, M., Pesenti, M., Costes, N., De Volder, A., & Seron, X. (2005). Task-independent semantic activation for numbers and animals. Brain research. *Cognitive brain research, 24,* 284–290.

Tonetti, L., Cellini, N., de Zambotti, M., Fabbri, M., Martoni, M., Fábregas, S. E., … Natale, V. (2013). Polysomnographic validation of a wireless dry headband technology for sleep monitoring in healthy young adults. P*hysiology & Behavior*, *118,* 185–188.

Tononi, G., & Cirelli, C. (2003). Sleep and synaptic homeostasis: a hypothesis. *Brain Research Bulletin, 62*, 143–150.

Tononi, G., & Cirelli, C. (2006). Sleep function and synaptic homeostasis. *Sleep Medicine Reviews, 10,* 49–62.

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. a, Wood, E. R., … Morris, R. G. M. (2007). Schemas and memory consolidation. *Science, 316,* 76–82.

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,* 891–895.

Tucker, M. a, & Fishbein, W. (2008). Enhancement of declarative memory performance following a daytime nap is contingent on strength of initial task acquisition. *Sleep, 31,* 197–203.

Tucker, M. a, 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,* 241–247.

Tzelgov, J. (1999). Automaticity and processing without awareness. *Psyche,5*(3), 18-23.

Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and Intentional Processing of Numerical Information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 166–179.

Van der Helm, E., Gujar, N., Nishida, M., & Walker, M. P. (2011). Sleep-dependent facilitation of episodic memory details. *PloS one, 6,* e27421.

Van Dongen, E. V, Thielen, J.-W., Takashima, A., Barth, M., & Fernández, G. (2012). Sleep supports selective retention of associative memories based on relevance for future utilization. *PloS one, 7,* e43426.

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*, 211–219.

Van Opstal, F., Gevers, W., De Moor, W., & Verguts, T. (2008). Dissecting the symbolic distance effect: Comparison and priming effects in numerical and nonnumerical orders. *Psychonomic Bulletin & Review, 15,* 419–425.

Vertes, R. P. (2004). Memory consolidation in sleep; dream or reality. *Neuron, 44,* 135–148.

Vertes, R. P., & Eastman, K. E. (2000). The case against memory consolidation in REM sleep. *The Behavioral and brain sciences, 23,* 867–876.

Walker, M. P. (2005). A refined model of sleep and the time course of memory formation. *The Behavioral and brain sciences, 28,* 51–64..

Walker, M. P. (2009). The role of sleep in cognition and emotion. *Annals of the New York Academy of Sciences*, *1156,* 168–197.

Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A., & Stickgold, R. (2002). Practice with sleep makes perfect: sleep-dependent motor skill learning. *Neuron, 35,* 205–211.

Walker, M. P., & Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron, 44,* 121–133.

Walker, M. P., & Stickgold, R. (2010). Overnight alchemy: sleep-dependent memory evolution. *Nature reviews. Neuroscience, 11(3),* 218.

Wechsler, D. (1999). *WASI manual.* San Antonio, Psychological Corporation.

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,* 1563–1569.

Wilkinson, G.S. (1993). *Wide Range Achievement Test 3– Administration manual.* Wilmington, DE: Jastak Associates, Inc.

Wilson, M., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science, 265,* 5–8.

Winocur, G., Moscovitch, M., & Bontempi, B. (2010). Memory formation and long-term retention in humans and animals: convergence towards a transformation account of hippocampal-neocortical interactions. *Neuropsychologia, 48,* 2339–2356.

Yap, M. J., Liow, S. J. R., Jalil, S. B., & Faizal, S. S. B. (2010). The Malay Lexicon Project: A database of lexical statistics for 9,592 words. *Behavior Research Methods, 42*, 992–1003.

Zamarian, L., Karner, E., & Benke, T. (2006). Knowing 7× 8, but not the meaning of “elephant”: Evidence for the dissociation between numerical and non-numerical semantic knowledge. *Neuropsychologia, 44,* 1708–1723.

Zhao, H., Chen, C., Zhang, H., Zhou, X., Mei, L., Chen, C., … Dong, Q. (2012). Is order the defining feature of magnitude representation? An ERP study on learning numerical magnitude and spatial order of artificial symbols. *PloS one*, *7*, e49565.

Zhou, X., Chen, Y., Chen, C., Jiang, T., Zhang, H., & Dong, Q. (2007). Chinese kindergartners’ automatic processing of numerical magnitude in stroop-like tasks. *Memory & Cognition, 35,* 464–470.