

# Markers of automaticity in sleep-associated consolidation of novel words



Elaine K.H. Tham<sup>\*1</sup>, Shane Lindsay<sup>2</sup>, M. Gareth Gaskell

Department of Psychology, University of York, York, YO10 5DD, United Kingdom

## ARTICLE INFO

### Article history:

Received 1 October 2014

Received in revised form

20 March 2015

Accepted 21 March 2015

Available online 26 March 2015

### Keywords:

Sleep

Memory consolidation

Integration

Automaticity

Word learning

## ABSTRACT

Two experiments investigated effects of sleep on consolidation and integration of novel form-meaning mappings using size congruity and semantic distance paradigms. Both paradigms have been used in previous studies to measure automatic access to word meanings. When participants compare semantic or physical font size of written word-pairs (e.g. BEE–COW), judgments are typically faster if relative sizes are congruent across both dimensions. Semantic distance effects are also found for well-established words, with semantic size judgments faster for pairs that differ substantially on this dimension. English-speaking participants learned novel form-meaning mappings with Mandarin (Experiment 1) or Malay (Experiment 2) words and were tested following overnight sleep or a similar duration awake. Judgements on English words controlled for circadian effects. The sleep group demonstrated selective stronger size congruity and semantic distance effects for novel word-pairs. This benefit occurred in Experiment 1 for semantic size comparisons of novel words, and in Experiment 2 on comparisons where novel pairs had large distances and font differences (for congruity effects) or in congruent trials (for semantic distance effects). Conversely, these effects were equivalent across sleep and wake for English words. Experiment 2 included polysomnography data and revealed that changes in the strength of semantic distance and congruity effects were positively correlated with slow-wave sleep and sleep spindles respectively. These findings support systems consolidation accounts of declarative learning and suggest that sleep plays an active role in integrating new words with existing knowledge, resulting in increased automatic access of the acquired knowledge.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

The idea that sleep benefits declarative learning stems from research on memory consolidation, which describes the process whereby a new memory representation becomes more resistant to forgetting over time (Stickgold and Walker, 2005). The active role of sleep in declarative memory consolidation is often explained by two-stage systems consolidation models (Frankland and Bontempi, 2005; McClelland, McNaughton, and O'Reilly, 1995), which incorporate both hippocampal and neocortical memory components. In these models, the hippocampal system encodes information swiftly, whereas the neocortical system is a slower learning, longer-term repository that gradually integrates new knowledge with existing long-term memories (Frankland and

Bontempi, 2005; O'Reilly and Norman, 2002). In an offline state, such as sleep, new memory representations are reactivated in the hippocampal system to promote neocortical storage. After multiple overnight cycles of offline reactivation, memory representations may become independent of the hippocampal system (Diekelmann and Born, 2010). Although there is debate over the aspects of sleep architecture involved in memory consolidation (Inostroza and Born, 2013), both slow-wave sleep (SWS) and sleep spindles (11–15 Hz activity in non-rapid eye movement (NREM) sleep) have been hypothesised to play an important role in the reactivation of memories in systems consolidation models (for reviews see Diekelmann and Born, 2010; Rasch and Born, 2013).

Although a link between sleep and the strengthening of new declarative memories is well established, few studies have examined the integration of new information into existing memory networks through the association between new and existing knowledge. In a novel-word learning experiment, Dumay and Gaskell (2007) provided evidence for the benefit of sleep on integration of new information with existing knowledge in a study that examined whether learning novel spoken word forms (e.g. “cathedruke”) would inhibit the recognition of familiar words (e.g.

\* Corresponding author.

E-mail address: [elaine\\_tham@sics.a-star.edu.sg](mailto:elaine_tham@sics.a-star.edu.sg) (E.K.H. Tham).

<sup>1</sup> Elaine K. H. Tham is now at the Agency for Science and Technology Research (A\*STAR), Singapore Institute for Clinical Sciences, Singapore

<sup>2</sup> Shane Lindsay is now at the Department of Psychology, University of Hull, United Kingdom

“cathedral”). Such an effect would suggest that the novel word had been integrated with existing neocortical knowledge, thus influencing the normal process of lexical competition in spoken word recognition (Gaskell and Dumay, 2003). This inhibitory effect on recognition of familiar words was absent soon after learning the novel forms, but emerged only after participants had slept overnight. These results suggest that the process of integrating new knowledge with existing knowledge is associated with sleep, in line with a systems consolidation account<sup>3</sup>. Using a similar paradigm, Tamminen et al. (2010) found that components of sleep may play specific roles in consolidation and integration of novel word forms. SWS was related to the strengthening of individual word memories in a speeded recognition memory test, whereas the inhibitory effect of the novel words on their existing neighbours during recognition was found to be correlated with spindle activity, suggesting that sleep spindles may play a crucial role in integrating novel words with existing lexical knowledge.

One way of thinking about the inhibitory effect discussed above is that it represents a change in the degree of *automaticity* involved in the access to the novel word form following sleep. Integration in the neocortex may enhance the automatic activation of the novel word when similar words are encountered. Automaticity is often thought of as a graded phenomenon, associated with a cluster of overlapping features. One description of such features is provided by Moors and De Houwer (2006), who analysed eight features characteristic of automatic processing: unintentional, purely stimulus driven, uncontrollable, autonomous, goal-independent, unconscious, efficient and fast. Depending on the nature of the task being investigated, these eight features may be present and relevant to different extents.

The motivation for the present work was to test directly whether sleep enhances aspects of automaticity in the access of newly learnt information as a result of integration. We examined putative evidence for automaticity in the context of access to word meaning. This might be expected as a consequence of systems consolidation (consolidation of new information from the hippocampal networks into existing neocortical networks) that results in a more direct route to accessing word meanings (Davis and Gaskell, 2009). Furthermore, while previous work has shown evidence of sleep affecting integration of novel forms (e.g. Dumay and Gaskell, 2007; Tamminen et al., 2010), here we seek to provide evidence for integration and/or automaticity in the context of form to meaning mappings.

Two paradigms that are argued to demonstrate aspects of automatic access to word meanings are size congruity and semantic distance effects. When given the task to select the larger of two written words presented next to each other (i.e. physical size of font) or the larger of the referents of the words (semantic size), correct judgments are faster when the relative sizes are congruent (e.g. BEE–COW) as opposed to incongruent (e.g. BEE–cow) along the physical and semantic dimensions (Paivio, 1975). This is known as the size congruity effect. In contrast, the semantic distance effect describes the result that response times (RTs) are faster for relative size judgements when the referents of the words have large size differences (e.g. BEE–COW) compared with referents that are closer (e.g. BEE–DOG) on the relevant dimension—in this case the size of the animal (Moyer and Landauer, 1967; Van Opstal et al., 2008). Rubinsten and Henik (2002) investigated both effects by showing participants animal word-pairs that differed in physical font size and semantic size. The semantic distance effect was found in semantic comparisons only, whereas the size congruity

effect was found in both physical and semantic comparison tasks. Rubinsten and Henik argued on the basis of these findings that meanings were rapidly and automatically activated during the word recognition process. Relating these two effects to the features of automaticity described above (Moors and De Houwer, 2006), both effects can be thought of as measures of efficiency and speed, as well as perhaps the level of controllability. The size congruity effect goes further in that it involves interference from an unattended variable, and so addresses features such as intentionality, goal-independence and autonomy. As such, the occurrence of the size congruity effect can be interpreted as revealing a greater level of sensitivity to automatic semantic access than the semantic distance effect (Rubinsten and Henik, 2002; Tzelgov et al., 1992; Tzelgov, 1999).

In the current paper, we used these properties of the size congruity and semantic distance effects to investigate the relationship between sleep and features of automaticity in the integration of new words into existing neocortical memory systems. We used a second language learning paradigm in which the novel forms share meanings with existing words in the participant's first language. Drawing from second language models such as the Revised Hierarchical Model, prior to integration, the relationship between the novel word forms and their meanings is hypothesised to be indirect and mediated by the translation from the existing word form (Kroll, Hell, Tokowicz, & Green, 2010) whereas it is plausible that after sleep-associated integration the access to the novel word's meaning is more direct (cf. Geukes et al., 2015). Another explanation relating to the complementary learning systems (CLS) account (Davis and Gaskell, 2009) of word learning suggests that when acquiring a second language (Lindsay and Gaskell, 2010), form-meaning cortical links for novel words are mediated by the hippocampal system during initial exposure. In this model it is predicted that after sleep-associated integration, there is a reduction in hippocampal mediation and a strengthening of more direct cortical links, resulting in more efficient and automatic access to the meanings.

## 2. Experiment 1

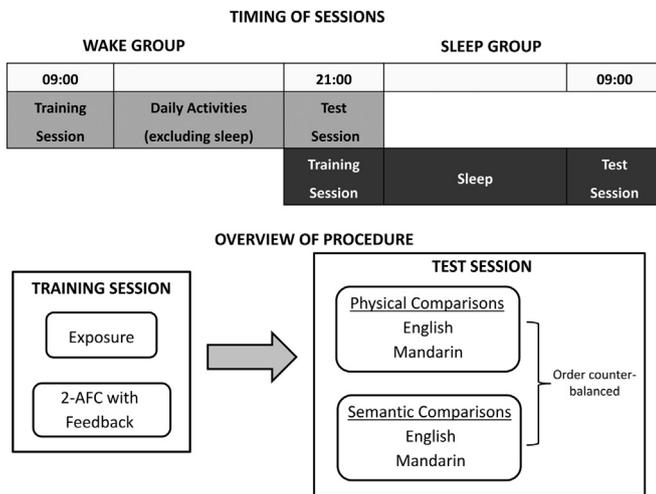
Using size congruity and semantic distance effects as hallmarks of established word representations, Experiment 1 explored the relationship between sleep and automaticity in novel word learning. Participants associated new Mandarin words with English animal names in the evening (sleep group) or morning (wake group), and were tested approximately 12 h later after a night of sleep or an equivalent time awake (see Fig. 1 for a summary of the design). It was hypothesised that if sleep benefits consolidation and integration of novel form-meaning pairings and more direct automatic access to those meanings, then participants who slept between learning and test would show greater semantic distance and size congruity effects than the wake group. Participants were also tested in the same way on established English words that were already fully consolidated in memory, in order to control for circadian confounds and fatigue. If effects found with the Mandarin words were due to sleep-associated consolidation, then we should not expect to find equivalent differences between the two groups of participants for the English words.

## 3. Method

### 3.1. Participants

Twenty-four monolingual native English speakers (9 males; mean age=21.7 years, range=18–31 years) with no known

<sup>3</sup> It should be noted that work by Lindsay and Gaskell (2013) suggests that integration can occur without sleep if new words are taught via spaced or interleaved learning.



**Fig. 1.** Overview of key methods and experimental procedure in [Experiment 1](#).

language or sleep disorders were recruited on a voluntary basis and gave informed consent before the experiment. Participants were randomly allocated to the wake ( $N=12$ ) or sleep ( $N=12$ ) group. They were instructed to have a night of normal sleep prior to the experiment and not to consume alcohol or recreational drugs throughout the experiment. The study was approved by the Department of Psychology Ethics Committee at the University of York.

### 3.2. Apparatus and materials

Stimuli were presented on a 15-inch screen using the DMDX experimental software ([Forster and Forster, 2003](#)) and responses were made with a USB-joypad. The main experimental stimuli consisted of 6 Mandarin characters and their corresponding meanings (see [Table A1](#) for a list of stimuli). In order to control for the complexity of the Mandarin characters, the learnt stimuli were not the actual names of each animal. Monosyllabic English animal names were selected from [Paivio \(1975\)](#) size norm ratings and divided into three groups representing small ( $M=1.5$ ,  $SD=0.5$ ), medium ( $M=3.4$ ,  $SD=0.6$ ) and large ( $M=6.3$ ,  $SD=1.3$ ) animals. Two sets of stimuli were used (Set A and B), where each set consisted of two small, two medium and two large animals.

All stimuli were presented in black against a white background, with Mandarin words presented in Chinese Black font and capitalised English stimuli in Consolas font. Colour photos selected from Google images depicted each animal. Experimental trials consisted of exposure, two-alternative forced choice (2-AFC) and test trials (see [Fig. 2](#)). In exposure trials a Mandarin word was centred on the upper half of the screen. Its corresponding animal name and photo were located in the bottom left and right quadrants respectively. For 2-AFC trials, if the target was a Mandarin word (centred on the upper half of the screen), the two choices were animal names (displayed in the bottom left and right quadrants) and if the target was an animal name, the two choices were Mandarin words. There were 120 feedback trials with all possible combinations of Mandarin words and animal names presented twice, counterbalancing items in the left and right quadrants.

For test trials, all valid pairings of stimuli were generated within the Mandarin-English sets. Stimuli were manipulated following [Rubinsten and Henik \(2002\)](#), where words in each pair varied on physical and semantic size. For the physical manipulation, words differed in height by either 1 mm (6 vs. 7 mm) or 4 mm (7 mm vs. 11 mm). For semantic size, words differed by either small (small-medium, medium-large animal) or large (small-

large animal) semantic distances. There were eight possible word-pairs differing by small semantic distances and four differing by large distances within each language set. To ensure that participants saw the same number of small and large semantic distance pairs, each large distance pair was presented twice. Congruity between the relative semantic and physical sizes was manipulated such that half the trials were congruent (the physically larger item was also semantically larger) and half incongruent (the physically larger item was semantically smaller). Each possible pair was then presented twice, with a centre-to-centre distance of 40 mm and counterbalancing screen location. Therefore, each participant saw test 256 trials for each language: 16 (small and large semantic distance pairs)  $\times$  2 (font height difference)  $\times$  2 (left/right counterbalancing)  $\times$  (congruity)  $\times$  2 comparison tasks (physical/semantic).

### 3.3. Procedure

For the wake group, training started at 9:00–10:30 a.m. and participants returned approximately 12 h later for the test session. Between sessions, wake participants carried on with their daily activities, and were asked to avoid napping and consuming caffeine or alcohol. For the sleep group, participants began training at 9:00–10:00 p.m. The subsequent 12-h interval included a night of normal sleep. To prevent effects of sleep inertia, sleep participants were tested at least 30 min after they awoke.

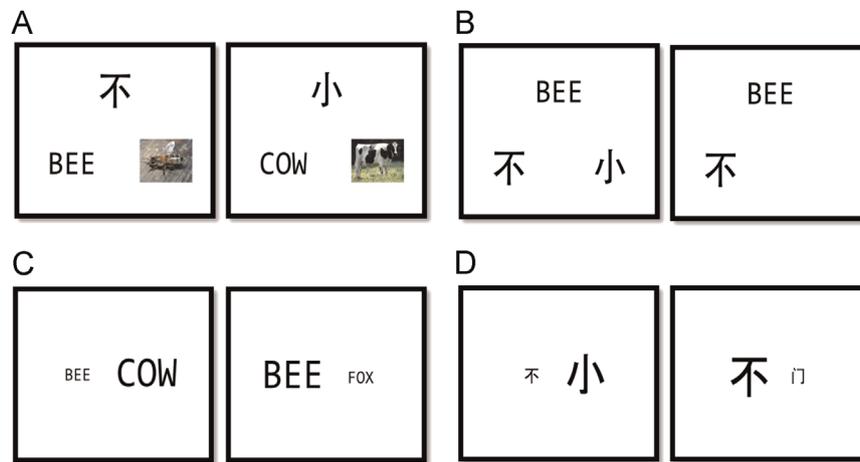
Participants were randomly allocated to Set A or Set B pairings. They were first given 18 exposure trials (3 blocks of 6) to associate each Mandarin character with an English animal name. Each trial lasted for 5000 ms and was followed by a 1000 ms inter-trial interval (blank screen). Next, participants completed a 2-AFC task where they were shown a target word and two ‘choice’ words. Participants had to select the alternative with the same meaning as the target. Trials were presented in a random order and displayed for 4000 ms or until participants responded. Participants were given feedback 500 ms after responding, whereby the incorrect ‘choice’ was removed.

For the test session, there were two types of comparison tasks (physical and semantic) run in a counterbalanced order. For each task, participants first viewed the English word-pairs consisting of animal names from the unlearned stimulus set, followed by the newly learnt Mandarin word-pairs. Participants were instructed to select the larger item as quickly and as accurately as possible. For physical comparisons, larger referred to the font height, whereas for semantic comparisons, larger referred to the size of the referent animal. For the English tasks, participants were given four practise trials. There were no practise trials for tasks in Mandarin. Before each trial, a fixation cross ‘+’ was displayed in the centre of the screen for 500 ms. Each trial was then presented for 4000 ms or until participants made a response.

## 4. Results

### 4.1. Training session

The overall mean error rate for the 2-AFC task was 2.7%. An ANOVA was performed on arcsine square root error rates, with stimulus set and group (sleep or wake) as between-participant factors. There was no significant difference in mean error rates based on stimulus set [ $F(1, 20)=1.66$ ,  $p=.21$ ] or time of learning [ $F(1, 20)=.74$ ,  $p=.40$ ]. These results suggest that learning was not influenced by the stimulus set participants were exposed to or circadian effects such as the time of day learning occurred.



**Fig. 2.** Examples of experimental trials: (A) Exposure, (B) 2 AFC with feedback, (C) test (control/English pairs, congruent-large semantic distance vs. incongruent-small semantic distance) and (D) test (novel word pairs, congruent-large semantic distance vs. incongruent-small semantic distance).

#### 4.2. Test session

RTs more than two standard deviations from a participant's mean for congruent or incongruent trials in each comparison task were excluded, resulting in overall data loss of 4.3%. Seven-way mixed ANOVAs were performed on the correct RTs (see Table 1) for the physical and semantic comparisons, with group, stimuli set and task order as between-participant factors, and language, congruity, physical (font) difference and semantic distance as within-participant factors. There was a significant main effect of language for both the semantic [ $F(1, 16)=16.41, p < .001$ ] and physical [ $F(1, 16)=15.63, p < .001$ ] comparison tasks, whereby participants were faster at responding to English than Mandarin word-pairs. This effect was not surprising, given the substantially greater familiarity of the English words. Given that responses were slower to the novel items overall, it is possible that likewise we should expect congruity and distance effects to be numerically greater for the novel than the familiar items. In order to ensure that results were not affected by the substantial difference in absolute RTs between languages, in cases where main effects or interactions were significant for the combined (English and Mandarin) ANOVA, further analysis was conducted to investigate if the main effects or interactions were present in each language separately. Interactions with stimuli sets are not reported (Pollatsek and Well, 1995), and interactions with order are not reported except for cases of theoretical interest. Results for physical comparisons tasks were not elaborated due to the lack of size congruity effects for the Mandarin word-pairs (see Appendix B).

#### 4.3. Semantic comparisons

Main effects and interactions involving group, congruity or semantic distance are detailed in Table C1. There was a marginal three-way interaction between congruity, group and language suggesting that group (wake vs. sleep) had more influence on the strength of the congruity effect for Mandarin than English words [ $F(1, 16)=3.57, p=.077$ ]. Crucially, the congruity  $\times$  group interaction was significant for Mandarin [ $F(1, 16)=7.87, p=.013$ ] but not English comparisons [ $F(1, 16)=.79, p=.39$ ] (see Fig. 3A). For the Mandarin stimuli, participants in the sleep group displayed clear size congruity effects [ $F(1, 8)=33.66, p < .001$ ], whereas there was no significant effect for the wake group [ $F(1, 8)=2.62, p=.14$ ]. In contrast, the size congruity effect was significant in both wake [ $F(1, 8)=7.27, p=.027$ ] and sleep [ $F(1, 8)=7.44, p=.026$ ] groups for English comparisons.

The semantic distance effect was significant in all semantic

**Table 1**

Mean RT in milliseconds for all correct responses in Experiment 1. Standard deviations are presented in parentheses.

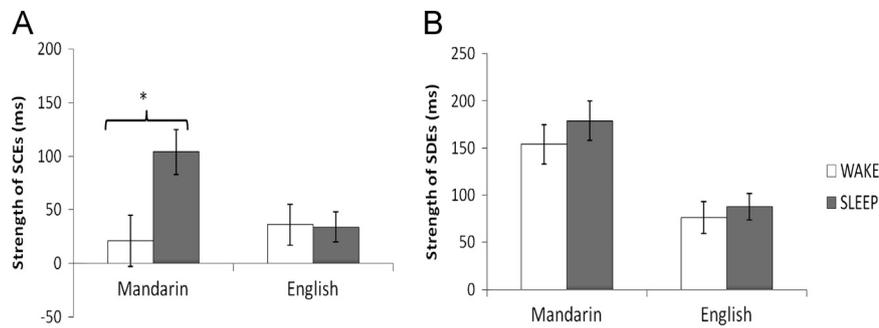
	English		Mandarin	
	Wake	Sleep	Wake	Sleep
Physical comparisons				
Size congruity				
Congruent	523 (87)	475 (87)	596 (117)	549 (138)
Incongruent	537 (83)	508 (111)	581 (142)	529 (114)
Strength	14 (7)	33 (35)	-15 (7)	-20 (24)
Cohen's d	.16	.33	.12	.16
Semantic distance				
Small	528 (87)	494 (97)	581 (107)	537 (121)
Large	532 (83)	489 (100)	597 (115)	540 (124)
Strength	-4 (3)	5 (3)	-16 (21)	-3 (7)
Cohen's d	.05	.05	.14	.02
Semantic comparisons				
Size congruity				
Congruent	922 (263)	863 (207)	1102 (290)	977 (197)
Incongruent	960 (287)	899 (201)	1133 (277)	1081 (190)
Strength	38 (21)	36 (7)	31 (35)	104 (24)
Cohen's d	.09	.12	.07	.37
Semantic distance				
Small	974 (280)	926 (218)	1195 (311)	1117 (207)
Large	908 (270)	836 (190)	1041 (249)	941 (180)
Strength	66 (10)	90 (21)	154 (35)	176 (48)
Cohen's d	.16	.26	.38	.93

Note: Strength=the difference scores between levels of congruity or semantic distance.

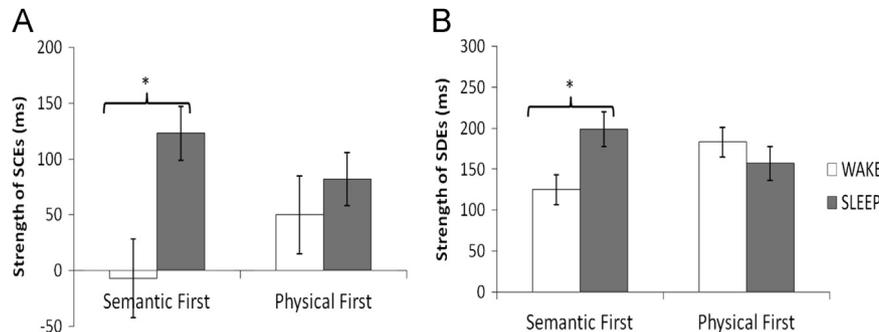
comparisons in the English [ $F(1, 16)=45.30, p < .001$ ] and Mandarin [ $F(1, 16)=143.69, p < .001$ ] word-pairs (see Fig. 3B). Unlike the congruity effect, there was no significant interaction between semantic distance, group and language [ $F(1, 16)=.001, p=1.00$ ].

There was also a significant four-way interaction between congruity, group, language and task order [ $F(1, 16)=4.94, p=.041$ ], apparently driven by task order effects (see Table C2). The congruity  $\times$  group interaction was present in the Mandarin word-pairs when the semantic comparison task was completed first [ $F(1, 8)=15.12, p < .01$ ], but not when it followed the physical comparison [ $F(1, 8)=.08, p=.79$ ] (Fig. 4A). Further analysis of Mandarin word-pair comparisons made when the semantic task was first revealed that the congruity effect was present in the sleep [ $F(1, 4)=29.10, p=.006$ ] but not the wake [ $F(1, 4)=.04, p=.86$ ] group. There were no equivalent interactions with task order for the English word-pair comparisons regardless of task order.

This pattern of results suggests that responses to the novel



**Fig. 3.** Automaticity effects in Experiment 1 for the semantic comparison task. (A) Size Congruity Effects and (B) Semantic Distance Effects. Bars are difference scores in RT (ms) between levels of size congruity or semantic distance. Error bars show standard error of the mean.



**Fig. 4.** Effects of task order (Semantic vs. Physical comparison first) for Mandarin word-pairs for semantic comparison in Experiment 1: (A) Size Congruity Effect, (B) Semantic Distance Effect. Bars are difference scores in RT (ms) between levels of size congruity or semantic distance. Error bars show standard error of the mean.

items may have altered during the test session due to the high level of exposure. Most likely the responses in the semantic task when it was presented first are the best indicator of the influence of the intervening period (sleep or wake). Interestingly, although the overall analysis suggested that wake and sleep groups were equivalent in terms of the strength of semantic distance effects, for participants who completed the semantic tasks first there was a significant interaction between semantic distance and group for the Mandarin word-pairs [ $F(1, 8) = 7.47, p = .026$ ] (see Fig. 4B). In these circumstances, the semantic distance effect was larger in the sleep than the wake group. In contrast, there were no equivalent effects for English words.

#### 4.4. Accuracy

Mean error rates for the English comparisons were 3.7% (wake) and 3.3% (sleep). For Mandarin comparisons, error rates were 4.7% (wake) and 4.4% (sleep). ANOVAs on arcsine square-root transformed errors showed no interaction with group.

## 5. Discussion

Experiment 1 investigated the relationship between sleep and representation of newly learnt words, using size congruity and semantic distance effects as measures of automaticity in form-meaning mappings. The main finding was that the size congruity effects for semantic comparisons of the newly learnt Mandarin words were found after sleep but not after an equivalent time awake. As hypothesised, there were no significant differences in the strength of these effects between wake and sleep groups for the English words. The lack of sleep-associated effects for English words suggests that the group difference during testing was not confounded by circadian effects or fatigue.

There were also significant interactions with task order. When participants made the physical judgement first, both groups

showed a significant influence of the physical properties of the stimuli on their semantic judgements, whereas when the semantic judgement was first this effect was only found for the sleep group. Order effects have emerged in previous studies using similar paradigms (Henik and Tzelgov, 1982), suggesting that prior experience with either judgement biases subsequent tasks. When participants were directed to make judgements about form, they found it difficult to avoid involving physical size in their subsequent semantic judgements. However, when the semantic judgement was first, wake participants processed meanings without any interference from physical size, whereas sleep participants automatically made use of both physical and semantic properties.

In contrast to the experimental hypothesis, there were no significant group differences for the distance effects in the semantic comparisons of Mandarin words. Both wake and sleep groups displayed significant semantic distance effects in the two languages. One explanation of this finding is, as discussed previously, that the size congruity effect may be a stronger demonstration of automatic semantic access than the semantic distance effect, given that the semantic comparison task does not require participants to attend to the interfering physical dimension (Rubinsten and Henik, 2002). The current findings suggest that a 12-h training-test interval was sufficient in eliciting the weaker demonstration of automaticity (semantic distance effects) for the new words, regardless of whether participants slept or not. Nonetheless, there was also a hint that sleep affected the strength of the distance effect. For the participants who carried out the semantic judgement first, the sleep group showed a bigger semantic distance effect than the wake group. However, in the absence of higher order interactions, this effect should be treated with caution.

Although participants were explicitly instructed to treat the new Mandarin words as a second language, it is important to bear in mind that the Mandarin words were non-alphabetic ideographs. Therefore, Mandarin words could potentially be viewed as less 'word-like' to monolingual English participants. The use of an alphabetic language that may be viewed as more word-like to native

English speaking participants would help to assess whether the changes in automaticity found for [Experiment 1](#) are limited to more imageable stimuli (cf. [Paivio, 1975](#)). More importantly, even though results from [Experiment 1](#) have provided initial support towards an association between sleep and knowledge integration, the argument would be stronger if we could show an association with particular components of sleep. Experiment 2 therefore examined whether individual sleep stages or components are associated with changes in congruity and semantic distance effects in second language word-pairs using polysomnography.

## 6. Experiment 2

Experiment 2 tested for an association between components of sleep and knowledge integration in word learning using Malay (alphabetic) words rather than Mandarin characters. Due to the apparent biasing effects of the physical comparisons on subsequent semantic judgements, Experiment 2 used only semantic comparisons. Participants in the sleep group had polysomnography data recorded during sleep. We predicted that the sleep group would display stronger size congruity and semantic distance effects than the wake group for Malay novel words but not for English control stimuli. Based on [Tamminen et al. \(2010\)](#), we further predicted that SWS and/or sleep spindle activity would play a facilitative role in the consolidation and strengthening of the novel form-meaning mappings leading to enhanced automaticity.

## 7. Method

### 7.1. Participants

Thirty-two monolingual native English speakers (13 males; mean age = 21.4 years, range: 18–34 years) who were non-smokers with no prior history of drug or alcohol abuse, were given a payment or course-credit for their participation. Participants had no reported neurological or psychiatric disorders, and no sleep disorders as assessed by the Pittsburgh Sleep Quality Index (PSQI; [Buysse et al., 1989](#)). Participants were randomly allocated to the wake ( $N=16$ ) or sleep ( $N=16$ ) experimental group. All remaining recruitment procedures were replicated from [Experiment 1](#).

### 7.2. Apparatus and materials

All apparatus and materials were similar to [Experiment 1](#), with the exception that [Experiment 2](#) used an extra three animal names in each set to increase statistical power. Animal names were translated into Malay to provide the novel stimuli. Therefore, the main experimental stimuli consisted of 18 English and 18 corresponding Malay animal names (see [Table A2](#) for the list of stimuli). It should be noted that there was no significant correlations between animal size ratings ([Paivio, 1975](#)) and word length,  $r = -.31$ ,  $p = .21$ .

The stimuli in the exposure, 2-AFC and test trials in Experiment 2 were manipulated similarly to those in [Experiment 1](#). As participants were trained on nine Malay animal names (instead of six in [Experiment 1](#)), there were nine exposure trials for each stimulus set. This also meant that participants saw more 2-AFC training trials in Experiment 2. There were six 2-AFC trials for each target animal, hence participants were exposed to 216 2-AFC training trials in total: 18 (9 English, 9 Malay) animals  $\times$  6 targets  $\times$  2 (left/right counterbalancing). As each animal could be both a target and a foil (i.e. incorrect 'choice' word), it should be noted each animal name (English/Malay) appeared with the same frequency: 12

times as the correct word and 12 times as foils. For the test session, word-pairs were manipulated similar to [Experiment 1](#), with 288 trials each language: 36 (small and large semantic size distance pairs)  $\times$  2 (physical height difference)  $\times$  2 (left/right counterbalancing)  $\times$  2 (congruity).

### 7.3. Procedure

Each participant underwent training and test sessions separated by approximately 10 h, with most of the procedural elements the same as in [Experiment 1](#). Before commencing the study, participants completed a consent form and reported their sleep quality via the PSQI. In order to evaluate alertness the Stanford Sleepiness Scale ([Hoddes et al., 1973](#)) was administered at the start of the training and test sessions. Participants in the wake group arrived in the laboratory at approximately 9:30 a.m. and returned for the test session at 7:30–7:45 p.m. For the sleep group, participants arrived at the laboratory at approximately 9 p.m. Prior to the training session, participants had electrodes attached for polysomnographic recording. The training session then began at approximately 10 p.m. and lasted for around 20 min. Participants then remained in the sleep laboratory for at least 8 h including time spent in overnight sleep.

### 7.4. Polysomnography recording

Polysomnographic data were recorded using an Embla N7000 system and Remlogix 3.0 software. A 10-channel montage was used following the international 10–20 system, with four scalp electrodes referenced to contralateral mastoids (C3-M2, C4-M1, O1-M2 and O2-M1), left and right electro-oculographic channels and 2 chin electromyographic channels. Technical specifications were as recommended in the [American Academy of Sleep Medicine Manual \(Iber, 2007\)](#). Data were scored manually by a single scorer in 30 s epochs according to [Rechtschaffen and Kales \(1968\)](#) criteria, but with Stages 3 and 4 scored collectively as SWS. Cross-scoring of 20% of the data by a second scorer revealed an inter-scoring agreement of 80%.

Spindle analysis was carried out by taking the mean activity from the bilateral central electrodes (C3 and C4) using the EEGlab toolbox ([Delorme and Makeig, 2004](#)) in Matlab (The Mathworks, Natick, MA, USA). NREM Stage 2 sleep and SWS were included in the analysis as the majority of sleep spindles occur in these stages. Epochs containing movements, arousal or noise artefacts were excluded. 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. \(2007\)](#).

## 8. Results

Data from one participant in the sleep group were excluded due to awakenings for more than 50% of the sleep period. Stanford Sleepiness Scale ratings were calculated for both training (wake:  $M=1.88$ ,  $SD=.72$ ; sleep:  $M=2.47$ ,  $SD=.91$ ) and test sessions (wake:  $M=2.25$ ,  $SD=1.12$ ; sleep:  $M=2.40$ ,  $SD=1.05$ ). PSQI ratings were also scored for the wake ( $M=4.06$ ,  $SD=1.39$ ) and sleep ( $M=3.60$ ,  $SD=2.09$ ) groups. There were no significant differences between wake and sleep groups on either rating at any session (all  $p$ 's  $> .27$ ).

### 8.1. Training session

The mean proportion of errors made for feedback trials was 3.8%. Similar to Experiment 1, there were no significant group

**Table 2**  
Mean RT in milliseconds for all correct responses in [Experiment 2](#). Standard deviations are presented in parentheses.

	English		Malay	
	Wake	Sleep	Wake	Sleep
Size congruity				
Congruent	904 (180)	918 (193)	1338 (328)	1157 (278)
Incongruent	912 (196)	946 (174)	1360 (328)	1187 (282)
Strength	8 (48)	28 (50)	22 (84)	30 (74)
Cohen's d	.03	.10	.04	.07
Semantic distance				
Small	953 (188)	983 (186)	1403 (340)	1237 (294)
Large	863 (188)	881 (182)	1294 (316)	1107 (267)
Strength	90 (48)	102 (46)	109 (84)	130 (58)
Cohen's d	.33	.39	.28	.32

Note: Strength=difference scores between levels of congruity or semantic distance.

difference in error rates based on stimulus set [ $F(1, 27)=.86, p=.36$ ] or time of training [ $F(1, 27)=.005, p=.94$ ].

## 8.2. Test session

### 8.2.1. Response times

RTs were excluded based on the same criteria as [Experiment 1](#), resulting in an average data loss of 1.6% per participant.

Mean RTs for all remaining correct responses were calculated and presented in [Table 2](#). A six-way mixed ANOVA was performed with group (wake vs. sleep) and stimuli set as between-participant factors, and language, congruity, physical size difference and semantic size distance as within-participant factors. There was a significant main effect of language with faster overall RTs in the English compared to the Malay words [ $F(1, 27)=70.27, p<.001$ ]. There was also a significant language x group interaction [ $F(1, 27)=5.85, p=.022$ ] whereby the sleep group had faster RTs than the wake group for the new Malay words but not for the English words. Main effects and interactions that included group and congruity effects or group and semantic distance effects are detailed in [Appendix D](#). As in [Experiment 1](#), in cases where main effects and interactions were significant for the combined (English and Malay) ANOVA, further analysis was conducted to investigate if the main effect and interactions were present in each language separately.

The semantic distance effect was significant across wake and sleep groups in both Malay [ $F(1, 27)=101.13, p<.001$ ] and English [ $F(1, 27)=193.23, p<.001$ ] comparisons, with RTs to pairs with large semantic distances faster than pairs with small distances. The size congruity effect was marginally significant for both the Malay [ $F(1, 27)=3.93, p=.058$ ] and English [ $F(1, 27)=2.93, p=.098$ ] comparisons. Unlike previous studies using the size congruity effect paradigm, the English items were not matched for word length, which may explain the lack of size congruity effects in English comparisons in [Experiment 2](#).

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 ([Table 3](#)). This suggests that group differences were more pronounced for the new Malay words than the existing English words. For the congruent trials (see [Fig. 5B](#)) there was a 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 [ $F(1, 27)=5.75, p=.024$ ]. In contrast, there was no equivalent significant

**Table 3**  
Mean RT in milliseconds for size congruity effects in trials with large semantic distances and large font differences; and semantic distance effects in congruent trials (trials that displayed sleep-associated automaticity effects). Standard deviations are presented in parentheses.

	English		Malay	
	Wake	Sleep	Wake	Sleep
Size congruity				
Congruent	835 (182)	850 (176)	1290 (330)	1076 (280)
Incongruent	861 (196)	872 (175)	1283 (321)	1152 (293)
Strength	26 (78)	22 (70)	-7 (106)	76 (107)
Cohen's d	.14	.13	.02	.27
Semantic Distance				
Small	955 (185)	968 (203)	1389 (341)	1228 (297)
Large	858 (186)	867 (189)	1282 (322)	1079 (264)
Strength	97 (58)	101 (51)	107 (100)	149 (59)
Cohen's d	.52	.51	.32	.53

Note: Strength=difference scores between levels of congruity or semantic distance

interaction for the incongruent trials [ $F(1, 27)=.19, p=.66$ ].

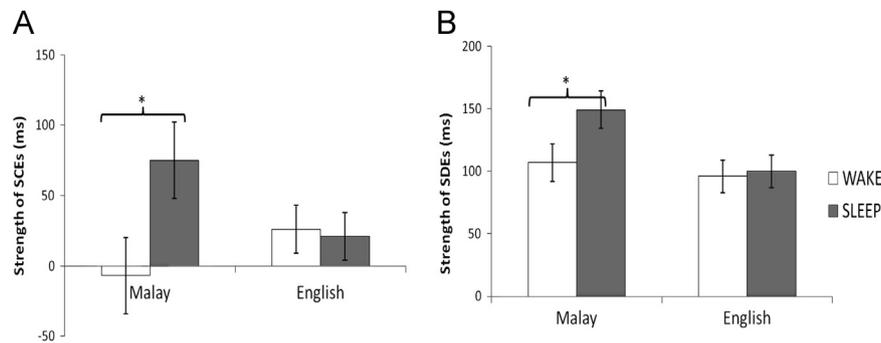
In addition, there was a marginal interaction between congruity, physical difference, semantic distance, group and language [ $F(1, 27)=3.10, p=.09$ ]. When considering the languages separately, the interaction between congruity, physical difference, semantic distance and group was marginally significant in the Malay [ $F(1, 27)=3.78, p=.062$ ] but not the English comparisons [ $F(1, 27)=.34, p=.56$ ], which suggests that the group differences are unique to the newly learnt Malay words ([Table 3](#)). Further analyses indicated that the latter interaction was driven by item-pairs which had both large semantic distances and physical font size differences ( $p$ -values were  $>.16$  for the other combinations of word-pairs). For these items, 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 examining the wake and sleep groups separately for the above trials with large semantic distances and physical differences, the sleep group displayed a significant size congruity effect [ $F(1, 13)=7.50, p=.017$ ], but the wake group did not [ $F(1, 14)=.27, p=.61$ ] (see [Fig. 5A](#)). All other main effects and interactions were non-significant except the interaction between congruity, physical difference, semantic distance for the English comparison [ $F(1, 27)=4.56, p=.042$ ], but there were no significant interactions when comparing individual variables.

## 8.3. Accuracy

For the English comparisons, the error rates were 3.5% (wake) and 3.9% (sleep). For the Malay comparisons, the error rates were 8.8% (wake) and 4.4% (sleep). An ANOVA was carried out on the accuracy data as in [Experiment 1](#). Similar to the RT analysis, there was a significant main effect of language where participants made more errors for the Malay words [ $F(1, 27)=7.11, p=.013$ ]. 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 English words [ $F(1, 14)=5.69, p=.023$ ] but there were no equivalent differences in error rates between the Malay and English words for the sleep group [ $F(1, 13)=.46, p=.51$ ].

## 8.4. Sleep stage and spindle analysis

Sleep stage data for the participants in the sleep group are given in [Table 4](#), with spindle activity data in [Table 5](#).



**Fig. 5.** Strength of automaticity effects in Experiment 2: (A) Size Congruity effect (for items with large semantic distances and large font differences), (B) Semantic Distance Effect (for congruent trials). Bars are difference scores in RT (ms) between levels of size congruity or semantic distance. Error bars show standard error of the mean.

**Table 4**

Sleep parameters for participants in the sleep group. Parentheses denote the standard deviation of the mean.

Sleep parameter	Mean time in minutes	Time as a percentage 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)

**Table 5**

Mean sleep spindle measures for central electrodes. Parentheses denote the standard deviation of the mean.

	C3	C4	Mean central (C3+C4)
Spindle Density (mean count per minute)	1.09 (0.55)	0.92 (0.41)	1.00 (0.46)
Total count of spindles	216 (188)	181 (88)	197 (100)

Correlations were calculated between measures that were sensitive to group differences (namely size congruity effect for all trials, size congruity effect for trials with large physical and semantic size difference, semantic distance effect for all trials and semantic distance effect for congruent trials) and the percentage of total sleep time spent in stage 2 sleep, SWS and REM sleep, as well as NREM spindle count (see Table 6). For each behavioural measure, Bonferroni correction was applied on the basis of the four sleep correlations tested, resulting in a *p*-value threshold of .0125.

When considering the overall size congruity effects, the correlation between NREM spindle count and the strength of the congruity effect across Malay comparisons was numerically strong but did not survive Bonferroni correction,  $r = .61, p = .015$ . However, there was a significant Bonferroni-corrected positive correlation between spindle count and strength of the size congruity effect in trials with large physical size difference and large semantic size distances for Malay comparisons,  $r = .71, p = .003$  but not for English comparisons,  $r = .19, p = .49$ . The Steiger's *r*-to-*z* transformation was applied to evaluate the significance of the difference between these two correlation coefficients (Steiger, 1980), with the difference approaching significance,  $z = 1.7, p = .089$ .

Analysis of semantic distance effects across all Malay trials with sleep stage and NREM spindle count showed no significant relationships. However, in the more focused analysis of semantic distance effect for congruent trials (which showed effects of wake/sleep group for the Malay trials), there was a significant

**Table 6**

Correlations between selected measures of automaticity in semantic comparisons 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)	<i>r</i>	0.10	−0.13	−0.006	<b>0.61</b>
		<i>p</i>	0.72	0.64	0.98	<b>0.015</b> †
	SCE (large physical and semantic size difference)	<i>r</i>	−0.21	0.005	0.05	<b>0.71</b>
		<i>p</i>	0.45	0.99	0.84	<b>0.003</b>
	SDE (all trials)	<i>r</i>	−0.42	0.51	−0.14	0.13
		<i>p</i>	0.11	0.05	0.62	0.63
SDE (congruent trials)	<i>r</i>	− <b>0.63</b>	<b>0.66</b>	−0.39	0.35	
	<i>p</i>	<b>0.013</b> †	<b>0.007</b>	0.15	0.19	
English comparisons	SCE (all trials)	<i>r</i>	0.34	−0.29	0.45	0.04
		<i>p</i>	0.21	0.29	0.08	0.86
	SCE (large physical and semantic size difference)	<i>r</i>	0.17	−0.23	0.45	0.19
		<i>p</i>	0.54	0.40	0.08	0.49
	SDE (all trials)	<i>r</i>	0.074	−0.14	0.43	−0.19
		<i>p</i>	0.79	0.60	0.10	0.48
SDE (congruent trials)	<i>r</i>	0.034	−0.053	0.26	−0.005	
	<i>p</i>	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

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,  $r = −.053, p = .85$ . In addition, Steiger's *Z* scores revealed a significant difference between the Malay and English correlation coefficients,  $z = 2.07, p = .038$ . There was also a negative correlation between the strength of the semantic distance effect for Malay congruent trials and time spent in stage 2 sleep,  $r = −.63, p = .013$ , but this did not survive Bonferroni correction. It appears likely that the correlation with stage 2 sleep was driven by a general negative correlation between time spent in stage 2 sleep and SWS,  $r = −.63, p < .001$ .

The two key correlations for newly learned items involved time in SWS (semantic distance effects for newly congruent trials) and spindle

count (congruity effects in trials with large physical and semantic differences between the items). To determine the independence of these effects partial correlations were conducted. When controlling for spindle count, the SWS correlation with Malay semantic distance effect (congruent) remained robust.  $r=.72$ ,  $p=.003$ . Likewise, when controlling for SWS, the correlation between spindle count and Malay size congruity effect (large physical difference and semantic distance) remained significant  $r=.71$ ,  $p=.004$ . The partial correlations revealed that SWS and NREM spindle count were uniquely related to semantic distance effect and size congruity effect respectively. It should be noted that when looking at the absolute rather than percentage time in stage 2, SWS and REM sleep, there was no change in the pattern of findings.

## 9. Discussion

In *Experiment 1*, participants who slept after learning showed stronger size congruity effects for newly learnt Mandarin words than those who remained awake for a comparable duration. *Experiment 2* supported the above finding as the sleep group displayed larger novel word size congruity effects than the wake group, but this time the effect of sleep was restricted to pairs with larger physical and semantic size differences. Similar to *Experiment 1*, when the wake and sleep groups were analysed separately, there was a significant size congruity effect in the sleep group but not the wake group for those items. In *Experiment 2*, we trained participants using Malay instead of Mandarin words, to increase the likelihood that participants saw the novel items as words from a new language. Unlike *Experiment 1* where there were no effects of sleep for semantic distance effect before taking into account task order effects, results from *Experiment 2* indicated that when making judgements of the new Malay words, participants who slept showed significantly larger semantic distance effects than wake participants for congruent trials. These results indicate that for participants who had just one night of sleep, access to the meanings of newly learnt words was more typical of integrated English words than those participants who remained awake, suggesting that sleep facilitated rapid activation of semantic information in the second language.

*Experiment 2* advanced upon *Experiment 1* by examining whether individual components of sleep had a specific role in enhancing automatic semantic access, as evidence of an active role of sleep on knowledge integration. 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 and semantic size differences. These findings suggest that individual components of sleep may also play an active role in enabling greater automaticity in processing new word meanings.

Sleep was also associated with faster and more accurate judgements of Malay words overall. There were no equivalent effects for the English word-pairs, indicating that the beneficial effect of sleep for the newly learnt words was not due to confounds such as circadian effects.

## 10. General discussion

Behavioural findings from both experiments indicate that sleep is associated with enhanced automaticity and integration of new form-meaning mappings with existing knowledge, whereby newly learnt words were more likely to exhibit properties similar to

established words after sleep than after wake. There were no equivalent differences between the sleep and wake groups in comparisons of existing English words, suggesting that the findings for the new words are not confounded by fatigue or time of day effects.

In both *Experiments 1* and *2*, size congruity effects for the newly learnt words were only exhibited when participants slept after learning. Relating the findings to the Revised Hierarchical Model of second language learning (Kroll et al., 2010), the absence of the size congruity effect for new words in the wake group suggests that even though participants in the wake group were able to access the new word meanings, the access is indirect and effortful – mediated by the translation from the existing word form. As we hypothesised, it is likely that after sleep-associated integration the access to the novel word's meaning is more direct, hence participants in the sleep group have displayed enhanced automaticity in processing new word meanings (see Geukes et al., 2015, for converging evidence based on the Stroop paradigm).

Our findings on the semantic distance and size congruity effects are also consistent with a complementary learning systems (CLS) consolidation model (Davis and Gaskell, 2009) applied to second language learning (Lindsay and Gaskell, 2010). The CLS account of word learning predicts that prior to sleep access to word meanings occur via hippocampal routes whereas after sleep-associated consolidation, the access is primarily reliant on neocortical routes. One potential consequence of increased reliance on neocortical memory is more automatic processing of the representations. Viewing this in relation to the features of automatic semantic access described in the Introduction (Moors and De Houwer, 2006), the semantic distance effect, which is thought to reflect a less direct measure of automaticity, was present in both the sleep and wake groups. However, the size congruity effect, which has more features of automaticity, 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 possible that sleep has a particular role to play in the transformation to a more unintentional, goal-independent and autonomous mode of processing.

By examining sleep polysomnography data, *Experiment 2* also examined whether there was a more direct relationship between the effects of sleep on integration in word learning. Based on Tamminen et al.'s (2010) word form learning study, where SWS was related to the consolidation and strengthening of novel word forms and spindle activity was related to the integration of the novel word form into the mental lexicon, it was predicted that both sleep spindle and SWS activity would play a 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 comparable trials involving well-consolidated English words. 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 and semantic size differences. Although further research is needed for a more comprehensive view, these findings give insight on how specific components of sleep might operate within the CLS model.

In *Experiment 1*, the semantic distance effect, a less direct measure of automatic semantic access, was found for the newly learnt words in both the wake and sleep groups. In contrast, the size congruity effect, a stronger measure of automatic semantic access, emerged only in the sleep group. Hence, it is plausible that the semantic distance effect would emerge with the strengthening

of links between word forms and meanings, even if these links are indirect and mediated, whereas the size congruity effect may benefit particularly from the strengthening of direct form-meaning links as a consequence of systems consolidation. Speculatively, SWS may play a main role in strengthening existing indirect form-meaning mappings thereby enhancing facets of automaticity such as speed and efficiency of processing. In comparison, sleep spindles and their interplay with slow oscillations and sharp-wave ripples (Buzsáki, 1989, 1996; Diekelmann and Born, 2010; Marshall and Born, 2007) may be more crucial for the hippocampal replay that facilitates the creation of direct cortical links between word forms and their meanings. This would enable processing to operate in a more unintentional and autonomous way via more direct access to shared meanings.

In addition to the task order effects found in [Experiment 1](#), there was also a lack of significant effects in the physical comparisons of the new Mandarin word-pairs, where participants were required to disregard the semantic size of the word pairs and selected the physically larger word based on font size. Participants were only given a night's consolidation period, whereas systems consolidation models have highlighted that consolidation occurs over multiple sleep-wake cycles rather than a single night (Diekelmann and Born, 2010; Frankland and Bontempi, 2005). Potentially, a fuller range of indicators of automaticity might be observable after a more extended sleep opportunity. A recent study by Tamminen and Gaskell (2013) supports this suggestion. They used a masked semantic priming paradigm to investigate automaticity of processing semantic information outside of participants' conscious awareness. 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 primed decisions to the existing words, it could be interpreted that the meanings of the novel words were integrated into the lexicon. When looking at individual test sessions, the main effect of priming was significant during the Day 8 test session but not for the other sessions (Day 1 and Day 2). Therefore, the changes in performance that we observe after a night's sleep in [Experiments 1](#) and [2](#) are likely to represent the beginning of a long and gradual process of consolidation and integration of the novel words and their meanings. We suspect that after a consolidation period spanning several days or more that is supplemented by additional training, automaticity effects may emerge for the physical comparison tasks.

In [Experiment 2](#), effects of sleep on automaticity emerged only in particular cases. For the semantic distance effect, sleep influences were only found for congruent trials (where the physical size and animal size matched). In their study of numerical cognition, Schwarz and Heinze (1998) found an effect of congruity on numerical distance effects in a number comparison task: when participants judged if a target digit was smaller or larger than a reference digit, the numerical distance effect was larger when the target-referent was congruent. Furthermore, when examining the English words in [Experiment 2](#), effect sizes for the semantic distance effects were larger for the congruent trials ([Table 3](#)) than across all trials ([Table 2](#)), adding to the above claim that the distance effect may sensitive to manipulations in congruency. As the distance effect seems more sensitive to congruent trials, it seems plausible that sleep-associated influences on semantic distance effects for newly learnt words would first emerge in congruent trials if participants only have one night's sleep for consolidation.

For the size congruity effect, sleep influences were only found in trials with large physical and semantic size differences. When considering the effects of physical font differences and semantic distances (in this case, numerical magnitude rather than animal

size) on size congruity, Cohen Kadosh and Henik (2006) found an additive effect of physical stimulus properties on size congruity in their study, where the strength of the size congruity effect increased with increased luminance differences between stimuli. This provides an analogue for the sleep-dependent benefit in size congruity effect for trials with larger physical differences in [Experiment 2](#). However, some studies of numerical cognition have found that the size congruity effect decreases with larger semantic (numerical) distance between items (Cohen Kadosh and Henik, 2006; Tzelgov et al., 1992). Nonetheless, the above studies used stimuli with established (well-learned) items, whereas the items learnt in [Experiment 2](#) only involved one night of consolidation. The case in our data that showed an association between sleep and enhanced size congruity effects can be thought of as the strongest congruity manipulation, with clearest discrepancy between relative semantic and physical sizes in the incongruent case. As above, we might expect the automaticity effect to be evident for a wider range of parameters as consolidation proceeds.

In conclusion, our experiments investigated whether sleep, and specific components of sleep, were associated with features of automaticity and the integration of new knowledge by comparing effects of size congruity and semantic distance effects for newly learnt words. Results suggested that 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. In terms of automaticity, one test of automaticity focusing on speed and efficiency of meaning access (the semantic distance effect) suggested that these components of automatic processing could be found after wake, but were somewhat enhanced after sleep. Meanwhile, the stricter test of automaticity (the size congruity effect) suggested that sleep facilitates the shift towards less intentional and more autonomous modes of processing. We see these changes as part of a much longer-term shift that enhances automatic aspects of word meaning retrieval. These findings contribute to a CLS model of word learning and extend our knowledge of the consequences of sleep-based memory consolidation, and accord with a view of consolidation in which there is a shift to greater speed, accuracy and automaticity in the access to new declarative memories over the course of sleep-wake cycles.

## Appendix A

See [Tables A1](#) and [A2](#).

**Table A1**  
List of stimuli for Experiment 1.

Mandarin character	Meaning*	Size group
<b>SET A</b>		
上	FLEA	Small
不	BEE	Small
又	DUCK	Medium
门	FOX	Medium
小	COW	Large
巴	MOOSE	Large
<b>SET B</b>		
巴	ANT	Small
又	SNAIL	Small
小	CAT	Medium
门	DOG	Medium
上	HORSE	Large
不	BEAR	Large

\* In order to control for the complexity of the Mandarin characters, the learnt stimuli were not the actual names of each animal.

**Table A2**  
List of Stimuli for Experiment 2.

Malay word	Meaning	Size group
<b>SET A</b>		
KUTU	ANT	Small
SEMUT	BEE	Small
SIPUT	FROG	Small
ITIK	DUCK	Medium
HELANG	EAGLE	Medium
KUCING	CAT	Medium
LEMBU	COW	Large
KUDA	HORSE	Large
RUSA	MOOSE	Large
<b>SET B</b>		
KUTU	FLEA	Small
SIPUT	SNAIL	Small
KETAM	CRAB	Small
RUBAH	FOX	Medium
BIRI	SHEEP	Medium
ANJING	DOG	Medium
BAGAL	MULE	Large
UNTA	CAMEL	Large
SINGA	LION	Large

**Table C1**  
ANOVA results of main effects and interactions with group and congruity or distance effects in semantic comparison tasks.

Factors	F	p
Language	16.41 <sup>***</sup>	< .001
Group (Wake/Sleep)	.71	.41
English	.40	.54
Mandarin	.81	.38
Congruity	87.81 <sup>***</sup>	< .001
English	14.68 <sup>***</sup>	< .001
Mandarin	26.55 <sup>***</sup>	< .001
Semantic distance	131.33 <sup>***</sup>	< .001
English	45.30 <sup>***</sup>	< .001
Mandarin	143.69 <sup>***</sup>	< .001
Physical font difference	.08	.78
English	1.29	.27
Mandarin	.72	.41
Order	.73	.40
English	2.17	.16
Mandarin	.012	.91
Congruity × Group × Language	3.57 <sup>#</sup>	.077
English (Congruity × Group)	.79	.39
Congruity (Sleep)	7.44 <sup>*</sup>	.026
Congruity (Wake)	7.27 <sup>*</sup>	.027
Mandarin (Congruity × Group)	7.87 <sup>*</sup>	.013
Congruity (Sleep)	33.66 <sup>***</sup>	< .001
Congruity (Wake)	2.62	.14
Congruity × Group × Order × Language	4.94 <sup>*</sup>	.041
Semantic distance × Group × Language	.001	1.00
Semantic distance × Group × Order × Language	.61	.45

\*  $p < .05$ .

\*\*\*  $p < .001$ .

#  $p < .1$ .

**Table C2**  
ANOVA interactions with group and congruity or distance effects based on task order.

Factors	F	p
<i>Semantic Comparison First</i>		
Congruity × Group × Language	8.99 <sup>*</sup>	.017
English (Congruity × Group)	.60	.45
Mandarin (Congruity × Group)	15.12 <sup>**</sup>	< .01
Semantic distance × Group × Language	.71	.42
English (Semantic distance × Group)	.01	.91
Mandarin (Semantic distance × Group)	7.47 <sup>*</sup>	.026
<i>Physical Comparison First</i>		
Congruity × Group × Language	.50	.49
English (Congruity × Group)	.61	.46
Mandarin (Congruity × Group)	.08	.79

\*  $p < .05$ .

\*\*  $p < .01$ .

**Table D1**  
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 <sup>***</sup>	< .001
Group	.71	.41
English	.24	.63
Malay	2.62	.12
Congruity (Cong)	7.49 <sup>*</sup>	.011
English	2.93 <sup>#</sup>	.098
Malay	3.93 <sup>#</sup>	.058
Semantic Distance (SD)	233.33 <sup>***</sup>	< .001
English	193.23 <sup>***</sup>	< .001
Malay	101.13 <sup>***</sup>	< .001
Physical Font Difference (PD)	.51	.48
English	1.17	.29
Malay	.063	.80
Cong × Group × Lang	.075	.78
English (Cong × Group)	.83	.36
Malay (Cong × Group)	.11	.74
SD × Group × Lang	.29	.59
English (SD × Group)	.51	.49
Malay (SD × Group)	1.04	.31
Cong × SD × Group × Lang	5.36 <sup>*</sup>	.028
English (Cong × SD × Group)	.27	.60
Malay (Cong × SD × Group)	5.93 <sup>*</sup>	.025
Congruent (SD × Group)	5.75 <sup>*</sup>	.024
Incongruent (SD × Group)	.19	.66
Cong × SD × PD × Group × Lang	3.10 <sup>#</sup>	.090
English (Cong × SD × PD × Group)	.34	.56
Malay (Cong × SD × PD × Group)	3.78 <sup>#</sup>	.062
Small SD	.47	.50
Small PD (Cong × Group)	1.03	.32
Large PD (Cong × Group)	2.04	.16
Large SD	4.83 <sup>*</sup>	.037
Small PD (Cong × Group)	.01	.92
Large PD (Cong × Group)	5.61 <sup>*</sup>	.025

\*  $p < .05$ .

\*\*\*  $p < .001$ .

#  $p < .1$ .

congruity effect was significant for the English [ $F(1, 16) = 9.47$ ,  $p < .01$ ], but not the Mandarin word-pairs [ $F(1, 16) = 2.44$ ,  $p = .14$ ], indicating a lack of automatic semantic access for the novel Mandarin words in the physical comparisons.

## Appendix C

See [Tables C1](#) and [C2](#).

## Appendix B

As we predicted following the results of [Rubinsten and Henik \(2002\)](#), a semantic distance effect was not found for physical comparisons for either Mandarin [ $F(1, 16) = .77$ ,  $p = .40$ ] or English [ $F(1, 16) = .005$ ,  $p = .95$ ] word-pairs. There was a significant interaction between language and congruity [ $F(1, 16) = 7.50$ ,  $p = .015$ ], with participants making swifter responses for congruent than incongruent trials for English but not Mandarin word-pairs. When looking at each language separately, the size

## Appendix D

See [Table D1](#).

## References

- American Academy of Sleep Medicine, & Iber, C. (2007). The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. American Academy of Sleep Medicine.
- 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 Res.* 28 (2), 193–213.
- Buzsáki, G., 1989. Two-stage model of memory trace formation: a role for noisy brain states. *Neuroscience* 31 (3), 551–570. doi:16/0306-4522(89)90423-5.
- Buzsáki, G., 1996. The hippocampo-neocortical dialogue. *Cereb. Cortex* 6 (2), 81–92.
- Cohen Kadosh, R., Henik, A., 2006. A common representation for semantic and physical properties. *Exp. Psychol. (formerly "Zeitschrift Für Experimentelle Psychologie")* 53 (2), 87–94. <http://dx.doi.org/10.1027/1618-3169.53.2.87>.
- Davis, M.H., Gaskell, M.G., 2009. A complementary systems account of word learning: neural and behavioural evidence. *Philos. Trans. R. Soc. Lond. Ser. B, Biol. Sci.* 364 (1536), 3773–3800. <http://dx.doi.org/10.1098/rstb.2009.0111>.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134 (1), 9–21. <http://dx.doi.org/10.1016/j.jneumeth.2003.10.009>.
- Diekelmann, S., Born, J., 2010. The memory function of sleep. *Nat. Rev. Neurosci.* 11 (2), 114–126. <http://dx.doi.org/10.1038/nrn2762>.
- Dumay, N., Gaskell, M.G., 2007. Sleep-associated changes in the mental representation of spoken words. *Psychol. Sci.* 18 (1), 35–39. <http://dx.doi.org/10.1111/j.1467-9280.2007.01845.x>.
- Ferrarelli, F., Huber, R., Peterson, M.J., Massimini, M., Murphy, M., Riedner, B., 2007. Reduced sleep spindle activity in schizophrenia patients. *Am. J. Psychiatry* 164 (3), 483–492. <http://dx.doi.org/10.1176/appi.ajp.164.3.483>.
- Forster, K.I., Forster, J.C., 2003. DMDX: a windows display program with millisecond accuracy. *Behav. Res. Methods, Instrum. Comput.* 35 (1), 116–124. Retrieved from.
- Frankland, P.W., Bontempi, B., 2005. The organization of recent and remote memories. *Nat. Rev. Neurosci.* 6 (2), 119–130. <http://dx.doi.org/10.1038/nrn1607>.
- Gaskell, M.G., Dumay, N., 2003. Lexical competition and the acquisition of novel words. *Cognition* 89 (2), 105–132. [http://dx.doi.org/10.1016/S0010-0277\(03\)00070-2](http://dx.doi.org/10.1016/S0010-0277(03)00070-2).
- Geukes, S., Gaskell, M.G., Zwitserlood, P., 2015. Stroop effects from newly learned color words: effects of memory consolidation and episodic context. *Front. Psychol.* <http://dx.doi.org/10.3389/fpsyg.2015.00278>
- Henik, A., Tzelgov, J., 1982. Is three greater than five: the relation between physical and semantic size in comparison tasks. *Mem. Cognit.* 10 (4), 389–395. Retrieved from.
- Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., Dement, W.C., 1973. Quantification of sleepiness: a new approach. *Psychophysiology* 10 (4), 431–436. <http://dx.doi.org/10.1111/j.1469-8986.1973.tb00801.x>.
- Inostroza, M., Born, J., 2013. Sleep for preserving and transforming episodic memory. *Ann. Rev. Neurosci.* April, 79–102. <http://dx.doi.org/10.1146/annurev-neuro-062012-170429>.
- Kroll, J., Hell, J., Van, Tokowicz, N., Green, D., 2010. The revised hierarchical model: a critical review and assessment. *Biling.: Lang. Cognit.* 13 (3), 373–381. <http://dx.doi.org/10.1017/S136672891000009X>.
- Lindsay, S., Gaskell, M.G., 2010. A complementary systems account of word learning in L1 and L2. *Lang. Learn.* 60, 45–63.
- Lindsay, S., Gaskell, M.G., 2013. Lexical integration of novel words without sleep. *Journal of Experimental Psychology. Learn. Mem. Cognit.* 39 (2), 608–622. <http://dx.doi.org/10.1037/a0029243>.
- Marshall, L., Born, J., 2007. The contribution of sleep to hippocampus-dependent memory consolidation. *Trends Cognit. Sci.* 11 (10), 442–450. <http://dx.doi.org/10.1016/j.tics.2007.09.001>.
- 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. *Psychol. Rev.* 102 (3), 419–457.
- Moors, A., De Houwer, J., 2006. Automaticity: a theoretical and conceptual analysis. *Psychol. Bull.* 132 (2), 297–326. <http://dx.doi.org/10.1037/0033-2909.132.2.297>.
- Moyer, R.S., Landauer, T.K., 1967. Time required for judgements of numerical inequality. *Nature* 215 (5109), 1519–1520. <http://dx.doi.org/10.1038/2151519a0>.
- O'Reilly, R.C., Norman, K.A., 2002. Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. *Trends Cognit. Sci.* 6 (12), 505–510.
- Paivio, A., 1975. Perceptual comparisons through the mind's eye. *Mem. Cognit.* 3 (6), 635–647. <http://dx.doi.org/10.3758/BF03198229>.
- Pollatsek, A., Well, A.D., 1995. On the use of counterbalanced designs in cognitive research: a suggestion for a better and more powerful analysis. *J. Exp. Psychol.: Learn. Mem. Cognit.* 21 (3), 785–794. <http://dx.doi.org/10.1037/0278-7393.21.3.785>.
- Rasch, B., Born, J., 2013. About sleep's role in memory. *Physiol. Rev.* 93 (2), 681–766. <http://dx.doi.org/10.1152/physrev.00032.2012>.
- 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.
- Rubinsten, O., Henik, A., 2002. Is an ant larger than a lion? *Acta Psychol.* 111 (1), 141–154.
- 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 (11), 1167–1179. Retrieved from.
- Steiger, J., 1980. Tests for comparing elements of a correlation matrix. *Psychol. Bull.* 245–251.
- Stickgold, R., Walker, M.P., 2005. Memory consolidation and reconsolidation: what is the role of sleep? *Trends Neurosci.* 28 (8), 408–415. doi:16/j.tins.2005.06.004.
- Tamminen, J., Gaskell, M.G., 2013. Novel word integration in the mental lexicon: evidence from unmasked and masked semantic priming. *Q. J. Exp. Psychol.* 66 (5), 1001–1025. <http://dx.doi.org/10.1080/17470218.2012.724694>.
- Tamminen, J., Payne, J., Stickgold, R., Wamsley, E., Gaskell, M.G., 2010. Sleep spindle activity is associated with the integration of new memories and existing knowledge. *J. Neurosci.* 30 (43), 14356–14360. <http://dx.doi.org/10.1523/JNEUROSCI.3028-10.2010>.
- Tzelgov, J., 1999. Automaticity and processing without awareness. *Psyche* 5, 18–23.
- Tzelgov, J., Meyer, J., Henik, A., 1992. Automatic and intentional processing of numerical information. *J. Exp. Psychol.: Learn. Mem. Cognit.* 18 (1), 166–179. Retrieved from.
- 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. *Psychon. Bull. Rev.* 15 (2), 419–425. <http://dx.doi.org/10.3758/PBR.15.2.419>.