

Running head: AUTOMATICITY IN CONSOLIDATION OF NOVEL WORDS

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Markers of Automaticity in Sleep-associated Consolidation of Novel Words

Elaine K. H. Tham¹, Shane Lindsay² and M. Gareth Gaskell

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

Corresponding Author: Elaine Kwang Hsia Tham

Telephone Number: (0065)63417423

Email: elaine_tham@sics.a-star.edu.sg

Author Note

¹Elaine K. H. Tham is now at the Singapore Institute for Clinical Sciences, Agency for Science and Technology Research (A*STAR), Brenner Centre for Molecular Medicine, Singapore.

²Shane Lindsay is now at the Department of Psychology, University of Hull, United Kingdom.

Markers of Automaticity in Sleep-associated Consolidation of Novel Words

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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 & Walker, 2005). The active role of sleep in declarative memory consolidation is often explained by two-stage systems consolidation models (Frankland & Bontempi, 2005; McClelland, McNaughton, & 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 & Bontempi, 2005; O'Reilly & 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 & Born, 2010). Although there is debate over the aspects of sleep architecture involved in memory consolidation (Inostroza & Born, 2013), both slow-wave sleep (SWS) and sleep spindles (11-15 Hz activity in non-rapid eye movement (NREM) sleep) have been hypothesized to play an important role in the reactivation of memories in systems consolidation models (for reviews see Diekelmann & Born, 2010; Rasch & 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. "cathedral"). Such an effect would suggest that the novel word had

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

been integrated with existing neocortical knowledge, thus influencing the normal process of lexical competition in spoken word recognition (Gaskell & 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 account¹. Using a similar paradigm, Tamminen, Payne, Stickgold, Wamsley and Gaskell (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

¹ 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.

1 examined putative evidence for automaticity in the context of access to word meaning. This
2 might be expected as a consequence of systems consolidation (consolidation of new
3 information from the hippocampal networks into existing neocortical networks) that results in
4 a more direct route to accessing word meanings (Davis & Gaskell, 2009). Furthermore, while
5 previous work has shown evidence of sleep affecting integration of novel forms (e.g. Dumay
6 & Gaskell, 2007; Tamminen et al., 2010), here we seek to provide evidence for integration
7 and/or automaticity in the context of form to meaning mappings.
8
9

10
11
12
13
14
15
16
17 Two paradigms that are argued to demonstrate aspects of automatic access to word
18 meanings are size congruity and semantic distance effects. When given the task to select the
19 larger of two written words presented next to each other (i.e. physical size of font) or the
20 larger of the referents of the words (semantic size), correct judgments are faster when the
21 relative sizes are congruent (e.g. BEE-COW) as opposed to incongruent (e.g. BEE-cow)
22 along the physical and semantic dimensions (Paivio, 1975). This is known as the size
23 congruity effect. In contrast, the semantic distance effect describes the result that response
24 times (RTs) are faster for relative size judgements when the referents of the words have large
25 size differences (e.g. BEE-COW) compared with referents that are closer (e.g. BEE-DOG) on
26 the relevant dimension - in this case the size of the animal (Moyer & Landauer, 1967; Van
27 Opstal, Gevers, De Moor, & Verguts, 2008). Rubinsten and Henik (2002) investigated both
28 effects by showing participants animal word-pairs that differed in physical font size and
29 semantic size. The semantic distance effect was found in semantic comparisons only, whereas
30 the size congruity effect was found in both physical and semantic comparison tasks.
31 Rubinsten and Henik argued on the basis of these findings that meanings were rapidly and
32 automatically activated during the word recognition process. Relating these two effects to the
33 features of automaticity described above (Moors and De Houwer, 2006), both effects can be
34 thought of as measures of efficiency and speed, as well as perhaps the level of controllability.
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 The size congruity effect goes further in that it involves interference from an unattended
2 variable, and so addresses features such as intentionality, goal-independence and autonomy.
3
4 As such, the occurrence of the size congruity effect can be interpreted as revealing a greater
5 level of sensitivity to automatic semantic access than the semantic distance effect (Rubinsten
6 & Henik, 2002; Tzelgov, Meyer, & Henik, 1992; Tzelgov, 1999).
7
8
9
10

11 In the current paper, we used these properties of the size congruity and semantic
12 distance effects to investigate the relationship between sleep and features of automaticity in
13 the integration of new words into existing neocortical memory systems. We used a second-
14 language learning paradigm in which the novel forms share meanings with existing words in
15 the participant's first language. Drawing from second-language models such as the Revised
16 Hierarchical Model, prior to integration, the relationship between the novel word forms and
17 their meanings is hypothesised to be indirect and mediated by the translation from the
18 existing word form (Kroll, Hell, Tokowicz, & Green, 2010) whereas it is plausible that after
19 sleep-associated integration the access to the novel word's meaning is more direct (cf.
20 Geukes, Gaskell, & Zwitserlood, 2015). Another explanation relating to the complementary
21 learning systems (CLS) account (Davis & Gaskell, 2009) of word learning suggests that when
22 acquiring a second language (Lindsay & Gaskell, 2010), form-meaning cortical links for
23 novel words are mediated by the hippocampal system during initial exposure. In this model it
24 is predicted that after sleep-associated integration, there is a reduction in hippocampal
25 mediation and a strengthening of more direct cortical links, resulting in more efficient and
26 automatic access to the meanings.
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

51 **Experiment 1**

52 Using size congruity and semantic distance effects as hallmarks of established
53 word representations, Experiment 1 explored the relationship between sleep and automaticity
54 in novel word learning. Participants associated new Mandarin words with English animal
55
56
57
58
59
60
61
62
63
64
65

names in the evening (sleep group) or morning (wake group), and were tested approximately 12 hours later after a night of sleep or an equivalent time awake (see Figure 1 for a summary of the design). It was hypothesized 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.

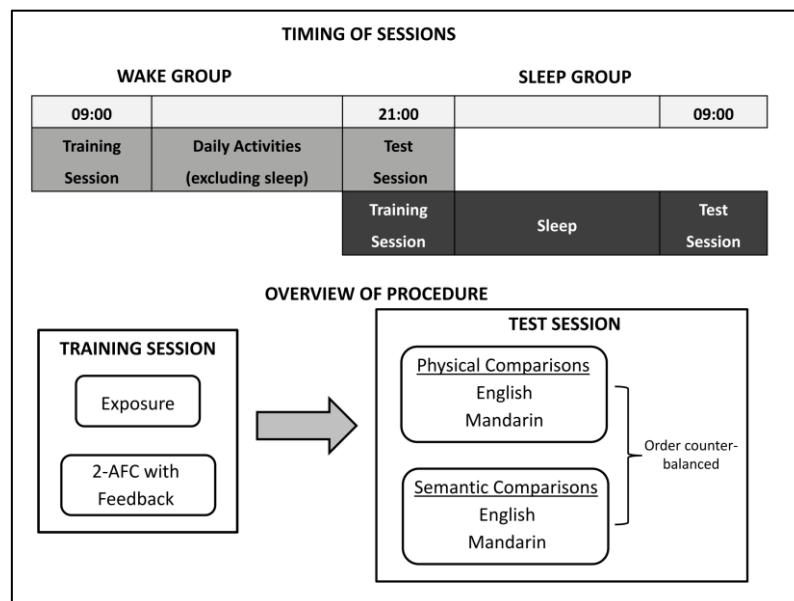


Figure 1. Overview of key methods and experimental procedure in Experiment 1

Method

Participants

Twenty-four monolingual native English speakers (9 males; mean age = 21.7 years, range = 18-31 years) with no known language or sleep disorders were recruited on a voluntary basis and gave informed consent before the experiment. Participants were

1 randomly allocated to the wake ($N = 12$) or sleep ($N = 12$) group. They were instructed to
2 have a night of normal sleep prior to the experiment and not to consume alcohol or
3 recreational drugs throughout the experiment. The study was approved by the Department of
4 Psychology Ethics Committee at the University of York.
5
6
7
8

9 **Apparatus and Materials**

10 Stimuli were presented on a 15-inch screen using the DMDX experimental software
11 (Forster & Forster, 2003) and responses were made with a USB-joypad. The main
12 experimental stimuli consisted of 6 Mandarin characters and their corresponding meanings
13 (see Table A1 for a list of stimuli). In order to control for the complexity of the Mandarin
14 characters, the learnt stimuli were not the actual names of each animal. Monosyllabic English
15 animal names were selected from Paivio's (1975) size norm ratings and divided into three
16 groups representing small ($M = 1.5$, $SD = 0.5$), medium ($M = 3.4$, $SD = 0.6$) and large ($M =$
17 6.3 , $SD = 1.3$) animals. Two sets of stimuli were used (Set A and B), where each set consisted
18 of two small, two medium and two large animals.
19
20
21
22
23
24
25
26
27
28
29
30
31
32

33 All stimuli were presented in black against a white background, with Mandarin words
34 presented in Chinese Black font and capitalized English stimuli in Consolas font. Colour
35 photos selected from Google images depicted each animal. Experimental trials consisted of
36 exposure, two-alternative forced choice (2-AFC) and test trials (see Figure 2). In exposure
37 trials a Mandarin word was centred on the upper half of the screen. Its corresponding animal
38 name and photo were located in the bottom left and right quadrants respectively. For 2-AFC
39 trials, if the target was a Mandarin word (centred on the upper half of the screen), the two
40 choices were animal names (displayed in the bottom left and right quadrants) and if the target
41 was an animal name, the two choices were Mandarin words. There were 120 feedback trials
42 with all possible combinations of Mandarin words and animal names presented twice,
43 counterbalancing items in the left and right quadrants.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

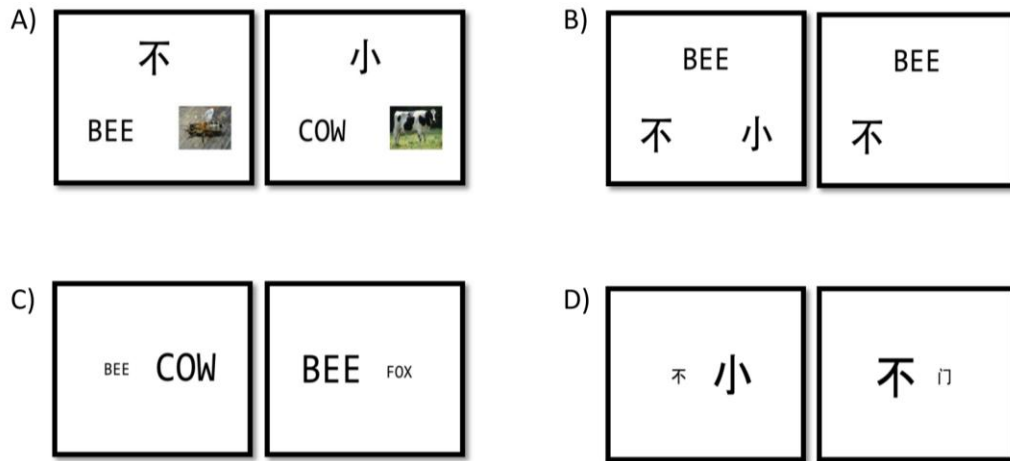


Figure 2. Examples of experimental trials: A) Exposure, B) 2 AFC with feedback, C) Test (control/English pairs, congruent-large semantic distance vs. incongruent-small semantic distance), D) Test (novel word pairs, congruent-large semantic distance vs. incongruent-small semantic distance).

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) x 2 (font height difference) x 2

(left/right counterbalancing) x 2 (congruity) x 2 comparison tasks (physical/semantic).

Procedure

For the wake group, training started at 9:00-10:30 a.m. and participants returned approximately 12 hours 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-hour interval included a night of normal sleep. To prevent effects of sleep inertia, sleep participants were tested at least 30 minutes 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 practice trials. There were no practice 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.

Results

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.

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)
<i>Strength</i>	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)
<i>Strength</i>	-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)
<i>Strength</i>	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)
<i>Strength</i>	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

Test Session

RTs more than two standard deviations from a participant's mean for congruent or

1 incongruent trials in each comparison task were excluded, resulting in overall data loss of
2 4.3%. Seven-way mixed ANOVAs were performed on the correct RTs (see Table 1) for the
3
4 physical and semantic comparisons, with group, stimuli set and task order as between-
5
6 participant factors, and language, congruity, physical (font) difference and semantic distance
7
8 as within-participant factors. There was a significant main effect of language for both the
9
10 semantic [$F(1, 16) = 16.41, p < .001$] and physical [$F(1, 16) = 15.63, p < .001$] comparison
11
12 tasks, whereby participants were faster at responding to English than Mandarin word-pairs.
13
14 This effect was not surprising, given the substantially greater familiarity of the English
15
16 words. Given that responses were slower to the novel items overall, it is possible that
17
18 likewise we should expect congruity and distance effects to be numerically greater for the
19
20 novel than the familiar items. In order to ensure that results were not affected by the
21
22 substantial difference in absolute RTs between languages, in cases where main effects or
23
24 interactions were significant for the combined (English and Mandarin) ANOVA, further
25
26 analysis was conducted to investigate if the main effects or interactions were present in each
27
28 language separately. Interactions with stimuli sets are not reported (Pollatsek & Well, 1995),
29
30 and interactions with order are not reported except for cases of theoretical interest. Results for
31
32 physical comparisons tasks were not elaborated due to the lack of size congruity effects for
33
34 the Mandarin word-pairs (see Appendix B).
35
36
37
38
39
40
41
42

43 **Semantic Comparisons**

44
45 Main effects and interactions involving group, congruity or semantic distance are
46
47 detailed in Table C1. There was a marginal three-way interaction between congruity, group
48
49 and language suggesting that group (wake vs. sleep) had more influence on the strength of the
50
51 congruity effect for Mandarin than English words [$F(1, 16) = 3.57, p = .077$]. Crucially, the
52
53 congruity x group interaction was significant for Mandarin [$F(1, 16) = 7.87, p = .013$] but not
54
55 English comparisons [$F(1, 16) = .79, p = .39$] (see Figure 3A). For the Mandarin stimuli,
56
57
58
59
60
61
62
63
64
65

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.

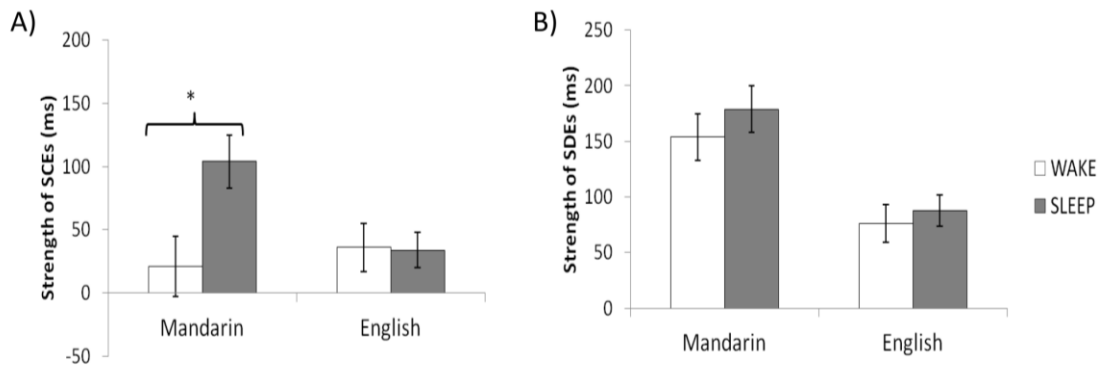


Figure 3. Automaticity effects in Experiment 1 for the semantic comparison task. A) Size Congruity Effects, B) Semantic Distance Effects. Bars are difference scores in RT (ms) between levels of size congruity or semantic distance. Error bars show standard error of the mean.

The semantic distance effect was significant in all semantic comparisons in the English [$F(1, 16) = 45.30, p < .001$] and Mandarin [$F(1, 16) = 143.69, p < .001$] word-pairs (see Figure 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 x 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 [$F(1, 8) = .08, p = .79$] (Figure 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 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 Figure 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.

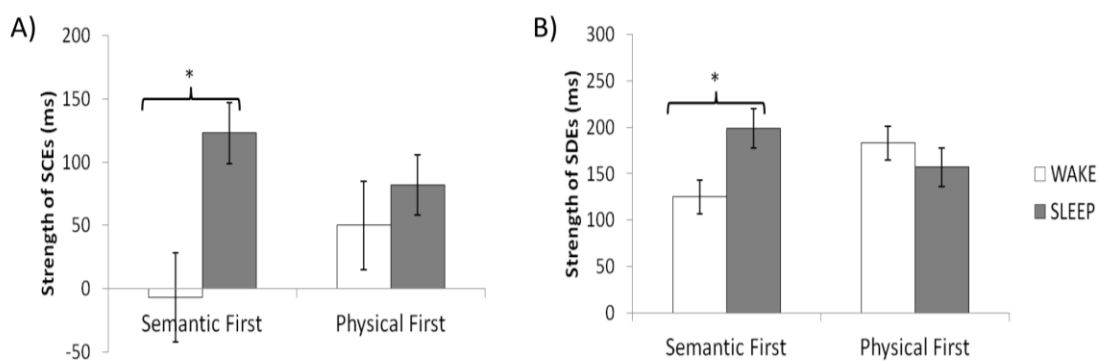


Figure 4. Effects of task order (Semantic vs. Physical comparison first) for Mandarin word-pairs for semantic comparison in Experiment 1: A) Size Congruity Effect, B) Semantic Distance Effect. Bars are difference scores in RT (ms) between levels of size congruity or semantic distance. Error bars show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

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.

Discussion

Experiment 1 investigated the relationship between sleep and representation of newly

1 learnt words, using size congruity and semantic distance effects as measures of automaticity
2 in form-meaning mappings. The main finding was that the size congruity effects for semantic
3 comparisons of the newly learnt Mandarin words were found after sleep but not after an
4 equivalent time awake. As hypothesized, there were no significant differences in the strength
5 of these effects between wake and sleep groups for the English words. The lack of sleep
6 associated effects for English words suggests that the group difference during testing was not
7 confounded by circadian effects or fatigue.
8
9

10
11
12
13
14
15
16
17 There were also significant interactions with task order. When participants made the
18 physical judgement first, both groups showed a significant influence of the physical
19 properties of the stimuli on their semantic judgements, whereas when the semantic judgement
20 was first this effect was only found for the sleep group. Order effects have emerged in
21 previous studies using similar paradigms (Henik & Tzelgov, 1982), suggesting that prior
22 experience with either judgement biases subsequent tasks. When participants were directed to
23 make judgements about form, they found it difficult to avoid involving physical size in their
24 subsequent semantic judgements. However, when the semantic judgement was first, wake
25 participants processed meanings without any interference from physical size, whereas sleep
26 participants automatically made use of both physical and semantic properties.
27
28
29
30
31
32
33
34
35
36
37
38
39
40

41 In contrast to the experimental hypothesis, there were no significant group differences
42 for the distance effects in the semantic comparisons of Mandarin words. Both wake and sleep
43 groups displayed significant semantic distance effects in the two languages. One explanation
44 of this finding is, as discussed previously, that the size congruity effect may be a stronger
45 demonstration of automatic semantic access than the semantic distance effect, given that the
46 semantic comparison task does not require participants to attend to the interfering physical
47 dimension (Rubinsten & Henik, 2002). The current findings suggest that a 12-hour training-
48 test interval was sufficient in eliciting the weaker demonstration of automaticity (semantic
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 distance effects) for the new words, regardless of whether participants slept or not.
2 Nonetheless, there was also a hint that sleep affected the strength of the distance effect. For
3 the participants who carried out the semantic judgement first, the sleep group showed a
4 bigger semantic distance effect than the wake group. However, in the absence of higher order
5 interactions, this effect should be treated with caution.
6
7
8
9
10

11 Although participants were explicitly instructed to treat the new Mandarin words as a
12 second language, it is important to bear in mind that the Mandarin words were non-alphabetic
13 ideographs. Therefore, Mandarin words could potentially be viewed as less ‘word-like’ to
14 monolingual English participants. The use of an alphabetic language that may be viewed as
15 more word-like to native English speaking participants would help to assess whether the
16 changes in automaticity found for Experiment 1 are limited to more imageable stimuli (cf.
17 Paivio, 1975). More importantly, even though results from Experiment 1 have provided
18 initial support towards an association between sleep and knowledge integration, the argument
19 would be stronger if we could show an association with particular components of sleep.
20 Experiment 2 therefore examined whether individual sleep stages or components are
21 associated with changes in congruity and semantic distance effects in second language word-
22 pairs using polysomnography.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

41 **Experiment 2**

42 Experiment 2 tested for an association between components of sleep and knowledge
43 integration in word learning using Malay (alphabetic) words rather than Mandarin characters.
44 Due to the apparent biasing effects of the physical comparisons on subsequent semantic
45 judgements, Experiment 2 used only semantic comparisons. Participants in the sleep group
46 had polysomnography data recorded during sleep. We predicted that the sleep group would
47 display stronger size congruity and semantic distance effects than the wake group for Malay
48 novel words but not for English control stimuli. Based on Tamminen et al. (2010), we further
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 predicted that SWS and/or sleep spindle activity would play a facilitative role in the
2 consolidation and strengthening of the novel form-meaning mappings leading to enhanced
3 automaticity.
4
5

6 7 **Method**

8 9 **Participants**

10 Thirty-two monolingual native English speakers (13 males; mean age = 21.4 years,
11 range: 18-34 years) who were non-smokers with no prior history of drug or alcohol abuse,
12 were given a payment or course-credit for their participation. Participants had no reported
13 neurological or psychiatric disorders, and no sleep disorders as assessed by the Pittsburgh
14 Sleep Quality Index (PSQI; Buysse, Reynolds, Monk, Berman, & Kupfer, 1989). Participants
15 were randomly allocated to the wake (N = 16) or sleep (N = 16) experimental group. All
16 remaining recruitment procedures were replicated from Experiment 1.
17
18
19
20
21
22
23
24
25
26
27

28 29 **Apparatus and Materials**

30 All apparatus and materials were similar to Experiment 1, with the exception that
31 Experiment 2 used an extra three animal names in each set to increase statistical power.
32 Animal names were translated into Malay to provide the novel stimuli. Therefore, the main
33 experimental stimuli consisted of 18 English and 18 corresponding Malay animal names (see
34 Table A2 for the list of stimuli). It should be noted that there was no significant correlations
35 between animal size ratings (Paivio, 1975) and word length, $r = -.31, p = .21$.
36
37
38
39
40
41
42
43
44
45

46 The stimuli in the exposure, 2-AFC and test trials in Experiment 2 were manipulated
47 similarly to those in Experiment 1. As participants were trained on nine Malay animal names
48 (instead of six in Experiment 1), there were nine exposure trials for each stimulus set. This
49 also meant that participants saw more 2-AFC training trials in Experiment 2. There were six
50 2-AFC trials for each target animal, hence participants were exposed to 216 2-AFC training
51 trials in total: 18 (9 English, 9 Malay) animals x 6 targets x 2 (left/right counterbalancing). As
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 each animal could be both a target and a foil (i.e. incorrect ‘choice’ word), it should be noted
2 each animal name (English/Malay) appeared with the same frequency: 12 times as the correct
3 word and 12 times as foils. For the test session, word-pairs were manipulated similar to
4 Experiment 1, with 288 trials each language: 36 (small and large semantic size distance-pairs)
5 x 2 (physical height difference) x 2 (left/right counterbalancing) x 2 (congruity).
6
7
8
9
10

11 **Procedure**

12 Each participant underwent training and test sessions separated by approximately 10
13 hours, with most of the procedural elements the same as in Experiment 1. Before
14 commencing the study, participants completed a consent form and reported their sleep quality
15 via the PSQI. In order to evaluate alertness the Stanford Sleepiness Scale (Hoddes, Zarcone,
16 Smythe, Phillips, & Dement, 1973) was administered at the start of the training and test
17 sessions. Participants in the wake group arrived in the laboratory at approximately 9:30 a.m.
18 and returned for the test session at 7:30-7:45 p.m. For the sleep group, participants arrived at
19 the laboratory at approximately 9 p.m. Prior to the training session, participants had
20 electrodes attached for polysomnographic recording. The training session then began at
21 approximately 10 p.m. and lasted for around 20 minutes. Participants then remained in the
22 sleep laboratory for at least 8 hours including time spent in overnight sleep.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

41 **Polysomnography recording**

42 Polysomnographic data were recorded using an Embla N7000 system and Remlogic
43 3.0 software. A 10-channel montage was used following the international 10-20 system, with
44 four scalp electrodes referenced to contralateral mastoids (C3-M2, C4-M1, O1-M2 and O2-
45 M1), left and right electro-oculographic channels and 2 chin electromyographic channels.
46 Technical specifications were as recommended in the American Academy of Sleep Medicine
47 Manual (Iber, 2007). Data were scored manually by a single scorer in 30-second epochs
48 according to Rechtschaffen and Kales (1968) criteria, but with Stages 3 and 4 scored
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

collectively as SWS. Cross-scoring of 20% of the data by a second scorer revealed an inter-scorer 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 & Makeig, 2004) in Matlab (Mathworks, 2011). 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).

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 $> .27$).

Training Session

The mean proportion of errors made for feedback trials was 3.8%. Similar to Experiment 1, there were no significant group difference in error rates based on stimulus set [$F(1, 27) = .86$, $p = .36$] or time of training [$F(1, 27) = .005$, $p = .94$].

Test Session

Response Times

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

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)
<i>Strength</i>	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)
<i>Strength</i>	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

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.

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)
<i>Strength</i>	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)
<i>Strength</i>	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

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

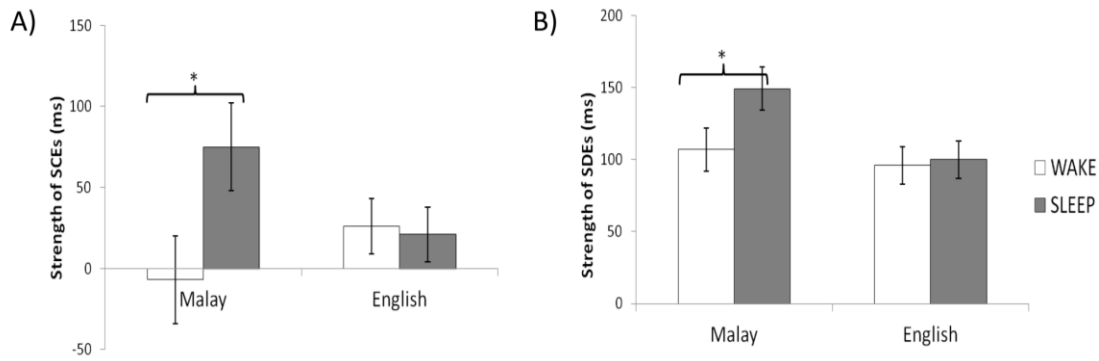


Figure 5. Strength of automaticity effects in Experiment 2: A) Size Congruity Effect (for items with large semantic distances and large font differences), B) Semantic Distance Effect (for congruent trials). Bars are difference scores in RT (ms) between levels of size congruity or semantic distance. Error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

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 Figure 5A). All other main effects and interactions were non-significant except the interaction between congruity, physical

1 difference, semantic distance for the English comparison [$F(1, 27) = 4.56, p = .042$], but there
 2 were no significant interactions when comparing individual variables.
 3

4 **Accuracy**

5 For the English comparisons, the errors rates were 3.5% (wake) and 3.9% (sleep). For
 6
 7 the Malay comparisons, the error rates were 8.8% (wake) and 4.4% (sleep). An ANOVA was
 8
 9 carried out on the accuracy data as in Experiment 1. Similar to the RT analysis, there was a
 10
 11 significant main effect of language where participants made more errors for the Malay words
 12
 13 [$F(1, 27) = 7.11, p = .013$]. There was also an interaction between language and group [$F(1,$
 14
 15 $27) = 4.23, p = .049$], where the wake group made significantly more errors for the Malay
 16
 17 compared to English words [$F(1, 14) = 5.69, p = .023$] but there were no equivalent
 18
 19 differences in error rates between the Malay and English words for the sleep group [$F(1, 13)$
 20
 21 $= .46, p = .51$].
 22
 23
 24
 25
 26
 27
 28

29 **Sleep stage and Spindle analysis**

30 Sleep stage data for the participants in the sleep group are given in Table 4, with
 31
 32 spindle activity data in Table 5.
 33
 34
 35
 36
 37
 38

39 Table 4. Sleep parameters for participants in the sleep group. Parentheses denote the standard
 40
 41 deviation of the mean.
 42

43 Sleep Parameter	44 Mean time in minutes	45 Time as a percentage of total sleep time
46 Total sleep time	486 (33)	
47 Wake time after sleep onset	18 (22)	
48 Sleep latency	25 (14)	
49 Stage 1	46 (21)	9.4 (4.1)
50 Stage 2	252 (22)	51.9 (5.0)
51 SWS	99 (26)	20.2 (4.8)
52 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.

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	0.61
		<i>p</i> 0.72	0.64	0.98	0.015 †
	SCE (large physical and semantic size difference)	<i>r</i> -0.21	0.005	0.05	0.71
		<i>p</i> 0.45	0.99	0.84	0.003
	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> -0.63	0.66	-0.39	0.35
		<i>p</i> 0.013 †	0.007	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

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 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

1 between the Malay and English correlation coefficients, $z = 2.07$, $p = .038$. There was also a
2 negative correlation between the strength of the semantic distance effect for Malay congruent
3 trials and time spent in stage 2 sleep, $r = -.63$, $p = .013$, but this did not survive Bonferroni
4 correction. It appears likely that the correlation with stage 2 sleep was driven by a general
5 negative correlation between time spent in stage 2 sleep and SWS, $r = -.63$, $p < .001$.
6
7
8
9
10

11 The two key correlations for newly learned items involved time in SWS (semantic
12 distance effects for congruent trials) and spindle count (congruity effects in trials with large
13 physical and semantic differences between the items). To determine the independence of
14 these effects partial correlations were conducted. When controlling for spindle count, the
15 SWS correlation with Malay semantic distance effect (congruent) remained robust. $r = .72$, p
16 $= .003$. Likewise, when controlling for SWS, the correlation between spindle count and
17 Malay size congruity effect (large physical difference and semantic distance) remained
18 significant $r = .71$, $p = .004$. The partial correlations revealed that SWS and NREM spindle
19 count were uniquely related to semantic distance effect and size congruity effect respectively.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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.

Discussion

66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

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

1 that participants saw the novel items as words from a new language. Unlike Experiment 1
2 where there were no effects of sleep for semantic distance effect before taking into account
3 task order effects, results from Experiment 2 indicated that when making judgements of the
4 new Malay words, participants who slept showed significantly larger semantic distance
5 effects than wake participants for congruent trials. These results indicate that for participants
6 who had just one night of sleep, access to the meanings of newly learnt words was more
7 typical of integrated English words than those participants who remained awake, suggesting
8 that sleep facilitated rapid activation of semantic information in the second language.
9

10
11
12
13
14
15
16
17
18
19 Experiment 2 advanced upon Experiment 1 by examining whether individual
20 components of sleep had a specific role in enhancing automatic semantic access, as evidence
21 of an active role of sleep on knowledge integration. When comparing behavioural findings
22 with sleep polysomnography data, the strength of the semantic distance effect for Malay
23 congruent trials was positively correlated with time spent in SWS, while the size congruity
24 effect was positively correlated with NREM spindle count for Malay word-pairs with larger
25 physical and semantic size differences. These findings suggest that individual components of
26 sleep may also play an active role in enabling greater automaticity in processing new word
27 meanings.
28
29
30
31
32
33
34
35
36
37
38
39
40

41 Sleep was also associated with faster and more accurate judgments of Malay words
42 overall. There were no equivalent effects for the English word-pairs, indicating that the
43 beneficial effect of sleep for the newly learnt words was not due to confounds such as
44 circadian effects.
45
46
47
48
49
50

51 **General Discussion**

52
53 Behavioural findings from both experiments indicate that sleep is associated with
54 enhanced automaticity and integration of new form-meaning mappings with existing
55 knowledge, whereby newly learnt words were more likely to exhibit properties similar to
56
57
58
59
60
61
62
63
64
65

1 established words after sleep than after wake. There were no equivalent differences between
2 the sleep and wake groups in comparisons of existing English words, suggesting that the
3 findings for the new words are not confounded by fatigue or time of day effects.
4
5

6
7 In both Experiment 1 and 2, size congruity effects for the newly learnt words were
8 only exhibited when participants slept after learning. Relating the findings to the Revised
9 Hierarchical Model of second language learning (Kroll et al., 2010), the absence of the size
10 congruity effect for new words in the wake group suggests that even though participants in
11 the wake group were able to access the new word meanings, the access is indirect and
12 effortful – mediated by the translation from the existing word form. As we hypothesized, it is
13 likely that after sleep-associated integration the access to the novel word's meaning is more
14 direct, hence participants in the sleep group have displayed enhanced automaticity in
15 processing new word meanings (see Geukes et al., 2015, for converging evidence based on
16 the Stroop paradigm).
17
18
19
20
21
22
23
24
25
26
27
28
29
30

31 Our findings on the semantic distance and size congruity effects are also consistent
32 with a complementary learning systems (CLS) consolidation model (Davis & Gaskell, 2009)
33 applied to second-language learning (Lindsay & Gaskell, 2010). The CLS account of word
34 learning predicts that prior to sleep access to novel word-meanings occur via hippocampal
35 routes whereas after sleep-associated consolidation, the access is primarily reliant on
36 neocortical routes. One potential consequence of increased reliance on neocortical memory is
37 more automatic processing of the representations. Viewing this in relation to the features of
38 automatic semantic access described in the Introduction (Moors & De Houwer, 2006), the
39 semantic distance effect, which is thought to reflect a less direct measure of automaticity, was
40 present in both the sleep and wake groups. However, the size congruity effect, which has
41 more features of automaticity, was present only in the sleep group for the new Malay words
42 (in the cases where interactions between group and congruity were present). Therefore it is
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 possible that sleep has a particular role to play in the transformation to a more unintentional,
2 goal-independent and autonomous mode of processing.
3

4
5 By examining sleep polysomnography data, Experiment 2 also examined whether
6
7 there was a more direct relationship between the effects of sleep on integration in word
8
9 learning. Based on Tamminen et al.'s (2010) word form learning study, where SWS was
10
11 related to the consolidation and strengthening of novel word forms and spindle activity was
12
13 related to the integration of the novel word form into the mental lexicon, it was predicted that
14
15 both sleep spindle and SWS activity would play a role in consolidation and integration of the
16
17 new Malay words. This prediction was supported, as participants who experienced greater
18
19 amounts of SWS also displayed stronger semantic distance effects in Malay word-pair
20
21 comparisons (congruent trials). Similar to the behavioural findings, there was no relationship
22
23 between SWS and the strength of the semantic distance effect for comparable trials involving
24
25 well-consolidated English words. In addition, there was a positive correlation between
26
27 NREM sleep spindle count and behavioural findings of the strength of the size congruity
28
29 effect for new Malay but not existing English word-pair comparisons with larger physical and
30
31 semantic size differences. Although further research is needed for a more comprehensive
32
33 view, these findings give insight on how specific components of sleep might operate within
34
35 the CLS model.
36
37
38
39
40
41
42

43
44 In Experiment 1, the semantic distance effect, a less direct measure of automatic
45
46 semantic access, was found for the newly learnt words in both the wake and sleep groups. In
47
48 contrast, the size congruity effect, a stronger measure of automatic semantic access, emerged
49
50 only in the sleep group. Hence, it is plausible that the semantic distance effect would emerge
51
52 with the strengthening of links between word forms and meanings, even if these links are
53
54 indirect and mediated, whereas the size congruity effect may benefit particularly from the
55
56 strengthening of direct form-meaning links as a consequence of systems consolidation.
57
58
59
60
61
62
63
64
65

1 Speculatively, SWS may play a main role in strengthening existing indirect form-meaning
2 mappings thereby enhancing facets of automaticity such as speed and efficiency of
3 processing. In comparison, sleep spindles and their interplay with slow oscillations and sharp-
4 wave ripples (Buzsáki, 1989, 1996; Diekelmann & Born, 2010; Marshall & Born, 2007) may
5 be more crucial for the hippocampal replay that facilitates the creation of direct cortical links
6 between word forms and their meanings. This would enable processing to operate in a more
7 unintentional and autonomous way via more direct access to shared meanings.
8
9

10
11
12
13
14
15
16
17 In addition to the task order effects found in Experiment 1, there was also a lack of
18 significant effects in the physical comparisons of the new Mandarin word-pairs, where
19 participants were required to disregard the semantic size of the word pairs and selected the
20 physically larger word based on font size. Participants were only given a night's
21 consolidation period, whereas systems consolidation models have highlighted that
22 consolidation occurs over multiple sleep-wake cycles rather than a single night (Diekelmann
23 & Born, 2010; Frankland & Bontempi, 2005). Potentially, a fuller range of indicators of
24 automaticity might be observable after a more extended sleep opportunity. A recent study by
25 Tamminen and Gaskell (2013) supports this suggestion. They used a masked semantic
26 priming paradigm to investigate automaticity of processing semantic information outside of
27 participants' conscious awareness. Participants were trained on the meanings of 34 novel
28 words and were tested on the same day (Day 1), after a night of consolidation (Day 2) and
29 after a week (Day 8). During the test session, the experimenters used the novel words as
30 primes for a lexical decision task of semantically related existing words. If the novel words
31 primed decisions to the existing words, it could be interpreted that the meanings of the novel
32 words were integrated into the lexicon. When looking at individual test sessions, the main
33 effect of priming was significant during the Day 8 test session but not for the other sessions
34 (Day 1 and Day 2). Therefore, the changes in performance that we observe after a night's
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 sleep in Experiments 1 and 2 are likely to represent the beginning of a long and gradual
2 process of consolidation and integration of the novel words and their meanings. We suspect
3 that after a consolidation period spanning several days or more that is supplemented by
4 additional training, automaticity effects may emerge for the physical comparison tasks.
5
6
7
8

9 In Experiment 2, effects of sleep on automaticity emerged only in particular cases.
10 For the semantic distance effect, sleep influences were only found for congruent trials (where
11 the physical size and animal size matched). In their study of numerical cognition, Schwarz
12 and Heinze (1998) found an effect of congruity on numerical distance effects in a number
13 comparison task: when participants judged if a target digit was smaller or larger than a
14 reference digit, the numerical distance effect was larger when the target-referent was
15 congruent. Furthermore, when examining the English words in Experiment 2, effect sizes for
16 the semantic distance effects were larger for the congruent trials (Table 3) than across all
17 trials (Table 2), adding to the above claim that the distance effect may sensitive to
18 manipulations in congruency. As the distance effect seems more sensitive to congruent trials,
19 it seems plausible that sleep-associated influences on semantic distance effects for newly
20 learnt words would first emerge in congruent trials if participants only have one night's sleep
21 for consolidation.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

41 For the size congruity effect, sleep influences were only found in trials with large
42 physical and semantic size differences. When considering the effects of physical font
43 differences and semantic distances (in this case, numerical magnitude rather than animal size)
44 on size congruity, Cohen Kadosh and Henik (2006) found an additive effect of physical
45 stimulus properties on size congruity in their study, where the strength of the size congruity
46 effect increased with increased luminance differences between stimuli. This provides an
47 analogue for the sleep-dependent benefit in size congruity effect for trials with larger physical
48 differences in Experiment 2. However, some studies of numerical cognition have found that
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 the size congruity effect decreases with larger semantic (numerical) distance between items
2 (Cohen Kadosh & Henik, 2006; Tzelgov et al., 1992). Nonetheless, the above studies used
3 stimuli with established (well-learnt) items, whereas the items learnt in Experiment 2 only
4 involved one night of consolidation. The case in our data that showed an association between
5 sleep and enhanced size congruity effects can be thought of as the strongest congruity
6 manipulation, with clearest discrepancy between relative semantic and physical sizes in the
7 incongruent case. As above, we might expect the automaticity effect to be evident for a wider
8 range of parameters as consolidation proceeds.
9

10
11
12
13
14
15
16
17
18
19 In conclusion, our experiments investigated whether sleep, and specific components
20 of sleep, were associated with features of automaticity and the integration of new knowledge
21 by comparing effects of size congruity and semantic distance effects for newly learnt words.
22 Results suggested that form-meaning mappings learnt during wakefulness are further
23 strengthened during sleep and that the integration of the individual mappings with existing
24 knowledge may be primarily sleep-dependent. In terms of automaticity, one test of
25 automaticity focusing on speed and efficiency of meaning access (the semantic distance
26 effect) suggested that these components of automatic processing could be found after wake,
27 but were somewhat enhanced after sleep. Meanwhile, the stricter test of automaticity (the size
28 congruity effect) suggested that sleep facilitates the shift towards less intentional and more
29 autonomous modes of processing. We see these changes as part of a much longer-term shift
30 that enhances automatic aspects of word meaning retrieval. These findings contribute to a
31 CLS model of word learning and extend our knowledge of the consequences of sleep-based
32 memory consolidation, and accord with a view of consolidation in which there is a shift to
33 greater speed, accuracy and automaticity in the access to new declarative memories over the
34 course of sleep-wake cycles.
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

References

- 1
2
3
4 Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The
5 Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research.
6 *Psychiatry Research*, 28(2), 193–213. Retrieved from
7 <http://www.ncbi.nlm.nih.gov/pubmed/2748771>
8
9
10 Buzsáki, G. (1989). Two-stage model of memory trace formation: A role for “noisy” brain
11 states. *Neuroscience*, 31(3), 551–570. doi:16/0306-4522(89)90423-5
12
13 Buzsáki, G. (1996). The hippocampo-neocortical dialogue. *Cerebral Cortex*, 6(2), 81–92.
14 Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8670641>
15
16
17 Cohen Kadosh, R., & Henik, A. (2006). A common representation for semantic and physical
18 properties. *Experimental Psychology (formerly “Zeitschrift Für Experimentelle*
19 *Psychologie”)*, 53(2), 87–94. doi:10.1027/1618-3169.53.2.87
20
21
22 Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning:
23 neural and behavioural evidence. *Philosophical Transactions of the Royal Society of*
24 *London. Series B, Biological Sciences*, 364(1536), 3773–3800.
25 doi:10.1098/rstb.2009.0111
26
27
28 Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-
29 trial EEG dynamics including independent component analysis. *Journal of Neuroscience*
30 *Methods*, 134(1), 9–21. doi:10.1016/j.jneumeth.2003.10.009
31
32
33 Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews.*
34 *Neuroscience*, 11(2), 114–126. doi:10.1038/nrn2762
35
36
37 Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation
38 of spoken words. *Psychological Science*, 18(1), 35–39. doi:10.1111/j.1467-
39 9280.2007.01845.x
40
41
42 Ferrarelli, F., Huber, R., Peterson, M. J., Massimini, M., Murphy, M., Riedner, B. a, ...
43 Tononi, G. (2007). Reduced sleep spindle activity in schizophrenia patients. *The*
44 *American Journal of Psychiatry*, 164(3), 483–492. doi:10.1176/appi.ajp.164.3.483
45
46
47 Forster, K. I., & Forster, J. C. (2003). DMDX: a windows display program with millisecond
48 accuracy. *Behavior Research Methods, Instruments, & Computers*, 35(1), 116–124.
49 Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12723786>
50
51
52 Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories.
53 *Nature Reviews Neuroscience*, 6(2), 119–130. doi:10.1038/nrn1607
54
55
56 Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words.
57 *Cognition*, 89(2), 105–132. doi:10.1016/S0010-0277(03)00070-2
58
59
60
61
62
63
64
65

- 1 Geukes, S., Gaskell, M. G., & Zwitserlood, P. (2015). Stroop effects from newly learned
2 color words : Effects of memory consolidation and episodic context. *Frontiers in*
3 *Psychology*. doi:10.3389/fpsyg.2015.00278
- 4
5 Henik, A., & Tzelgov, J. (1982). Is three greater than five: the relation between physical and
6 semantic size in comparison tasks. *Memory & Cognition*, *10*(4), 389–395. Retrieved
7 from <http://www.ncbi.nlm.nih.gov/pubmed/7132716>
8
- 9
10 Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of
11 sleepiness: A new approach. *Psychophysiology*, *10*(4), 431–436. doi:10.1111/j.1469-
12 8986.1973.tb00801.x
- 13
14 Inostroza, M., & Born, J. (2013). Sleep for preserving and transforming episodic memory.
15 *Annual Review of Neuroscience*, (April), 79–102. doi:10.1146/annurev-neuro-062012-
16 170429
- 17
18
19 Kroll, J., Hell, J. Van, Tokowicz, N., & Green, D. (2010). The Revised Hierarchical Model:
20 A critical review and assessment. *Bilingualism: Language and Cognition*, *13*(3), 373–
21 381. doi:10.1017/S136672891000009X.The
- 22
23
24 Lindsay, S., & Gaskell, M. G. (2010). A complementary systems account of word learning in
25 L1 and L2. *Language Learning*, *60*, 45–63.
- 26
27
28 Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep.
29 *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *39*(2), 608–22.
30 doi:10.1037/a0029243
- 31
32
33 Marshall, L., & Born, J. (2007). The contribution of sleep to hippocampus-dependent
34 memory consolidation. *Trends in Cognitive Sciences*, *11*(10), 442–450.
35 doi:10.1016/j.tics.2007.09.001
- 36
37
38 McClelland, J., McNaughton, B., & O'Reilly, R. (1995). Why there are complementary
39 learning systems in the hippocampus and neocortex: insights from the successes and
40 failures of connectionist models of learning and memory. *Psychological Review*, *102*(3),
41 419–457. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7624455>
42
- 43
44 Moors, A., & De Houwer, J. (2006). Automaticity: a theoretical and conceptual analysis.
45 *Psychological Bulletin*, *132*(2), 297–326. doi:10.1037/0033-2909.132.2.297
- 46
47
48 Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical
49 inequality. *Nature*, *215*(5109), 1519–1520. doi:10.1038/2151519a0
- 50
51
52 O'Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to
53 memory: Advances in the complementary learning systems framework. *Trends in*
54 *Cognitive Sciences*, *6*(12), 505–510.
- 55
56
57 Paivio, A. (1975). Perceptual comparisons through the mind's eye. *Memory & Cognition*,
58 *3*(6), 635–647. doi:10.3758/BF03198229
- 59
60
61
62
63
64
65

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Reviews*, 93(2), 681–766. doi:10.1152/physrev.00032.2012
- Rubinsten, O., & Henik, A. (2002). Is an ant larger than a lion? *Acta Psychologica*, 111(1), 141–154. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12102118>
- 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 <http://www.ncbi.nlm.nih.gov/pubmed/9842762>
- Steiger, J. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 245–251.
- Stickgold, R., & Walker, M. P. (2005). Memory consolidation and reconsolidation: what is the role of sleep? *Trends in Neurosciences*, 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. *Quarterly Journal of Experimental Psychology (2006)*, 66(5), 1001–1025. doi: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. *The Journal of Neuroscience*, 30(43), 14356–14360. doi: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. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(1), 166–179. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0278739302010352>
- 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(2), 419–425. doi:10.3758/PBR.15.2.419

Appendix A

Table A1. List of Stimuli for Experiment 1

Mandarin Character	Meaning*	Size Group
SET A		
上	FLEA	Small
不	BEE	Small
又	DUCK	Medium
门	FOX	Medium
小	COW	Large
巴	MOOSE	Large
SET 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
SET A		
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
SET 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

Appendix B

1
2
3
4 As we predicted following the results of Rubinsten and Henik (2002), a semantic distance
5
6 effect was not found for physical comparisons for either Mandarin [$F(1, 16) = .77, p = .40$] or
7
8 English [$F(1, 16) = .005, p = .95$] word-pairs. There was a significant interaction between
9
10 language and congruity [$F(1, 16) = 7.50, p = .015$], with participants making swifter
11
12 responses for congruent than incongruent trials for English but not Mandarin word-pairs.
13
14
15 When looking at each language separately, the size congruity effect was significant for the
16
17 English [$F(1, 16) = 9.47, p < .01$], but not the Mandarin word-pairs [$F(1, 16) = 2.44, p = .14$],
18
19 indicating a lack of automatic semantic access for the novel Mandarin words in the physical
20
21 comparisons.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Appendix C

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

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

Note: # $p < .1$; * $p < .05$; *** $p < .001$

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

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

* $p < .05$; ** $p < .01$.

Appendix D

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

Factors	<i>F</i>	<i>p</i>
Language (Lang)	70.27***	< .001
Group	.71	.41
English	.24	.63
Malay	2.62	.12
Congruity (Cong)	7.49*	.011
English	2.93#	.098
Malay	3.93#	.058
Semantic Distance (SD)	233.33***	< .001
English	193.23***	< .001
Malay	101.13***	< .001
Physical Font Difference (PD)	.51	.48
English	1.17	.29
Malay	.063	.80
Cong x Group x Lang	.075	.78
English (Cong x Group)	.83	.36
Malay (Cong x Group)	.11	.74
SD x Group x Lang	.29	.59
English (SD x Group)	.51	.49
Malay (SD x Group)	1.04	.31
Cong x SD x Group x Lang	5.36*	.028
English (Cong x SD x Group)	.27	.60
Malay (Cong x SD x Group)	5.93*	.025
Congruent (SD x Group)	5.75*	.024
Incongruent (SD x Group)	.19	.66
Cong x SD x PD x Group x Lang	3.10#	.090
English (Cong x SD x PD x Group)	.34	.56
Malay (Cong x SD x PD x Group)	3.78#	.062
Small SD	.47	.50
Small PD (Cong x Group)	1.03	.32
Large PD (Cong x Group)	2.04	.16
Large SD	4.83*	.037
Small PD (Cong x Group)	.01	.92
Large PD (Cong x Group)	5.61*	.025

#*p* < .1; **p* < .05; ****p* < .001

TIMING OF SESSIONS

WAKE GROUP		SLEEP GROUP		
09:00		21:00		09:00
Training Session	Daily Activities (excluding sleep)	Test Session		
		Training Session	Sleep	Test Session

OVERVIEW OF PROCEDURE

