Vast amounts of encoded items nullify but do not reverse the effect of sleep on declarative memory

1 **Luca D. Kolibius[1], Jan Born[2,3], Gordon B. Feld[2,4*]**

2 [1] University of Birmingham, School of Psychology, Centre for Human Brain Health

3 [2] University of Tübingen, Institute of Medical Psychology and Behavioral Neurobiology

4 [3] University of Tübingen, Centre for Integrative Neuroscience

5 [4] Central Institute of Mental Health, Mannheim, Germany

6 *** Correspondence:**
7 Gordon B. Feld
8 gordon.feld@zi-mannheim.de

Abstract

12 Sleep strengthens memories by repeatedly reactivating associated neuron ensembles. Our
13 studies show that although long-term memory for a medium number of word-pairs (160)
14 benefits from sleep, a large number (320) does not. This suggest an upper limit to the amount
15 of information that has access to sleep-dependent declarative memory consolidation, which is
16 possibly linked to the availability of reactivation opportunities. Due to competing processes of
17 global forgetting that are active during sleep, we hypothesised that even larger amounts of
18 information would enhance the proportion of information that is actively forgotten during
19 sleep. In the present study, we aimed to induce such forgetting by challenging the sleeping
20 brain with vast amounts of to be remembered information. For this, 80 participants learned a
21 very large number of 640 word-pairs over the course of an entire day and then either slept or
22 stayed awake during the night. Recall was tested after another night of regular sleep. Results
23 revealed comparable retention rates between the sleep and wake groups. Although this null-
24 effect can be reconciled the concept of limited capacities available for sleep-dependent
25 consolidation, it contradicts our hypothesis that sleep would increase forgetting compared to
26 the wake group. Additional exploratory analyses relying on equivalence testing and Bayesian
27 statistics reveal that there is evidence against sleep having a detrimental effect on the retention
28 of declarative memory at high information loads. We argue that forgetting occurs in both
29 wake and sleep states through different mechanisms, i.e., through increased interference and
30 through global synaptic downscaling, respectively. Both of these processes might scale
31 similarly with information load.

1   Introduction

33 It is undisputed that sleep is integral to the formation of long-term memory (Diekelmann &
34 Born, 2010; Klinzing et al., 2019; Rasch & Born 2013; Walker & Stickgold, 2010). Initially,
35 the idea prevailed that sleep predominantly acts as a passive shield against interference from
36 novel information, as put forward by Jenkins and Dallenbach (Jenkins & Dallenbach, 1924).
37 Even though modern interpretations of this framework still exist, it is now generally accepted

38 that sleep plays an active role for memory (Ellenbogen, Payne, & Stickgold, 2006), with the
39 two-stage model of memory formation (Diekelmann & Born, 2010; Marr, 1971; Klinzing, et
40 al., 2019; Rasch & Born, 2013) being the prevailing model used in declarative memory
41 research. First introduced by Marr (1971) it offers a solution to the "stability-plasticity-
42 dilemma" (Abraham & Robins, 2005), which refers to the problem how a system can learn
43 new information rapidly and in succession without overwriting older memories (Robins,
44 1995).
45   Initially the hippocampus binds together distributed information in the cortex during
46 encoding (Battaglia et al., 2011). During subsequent sleep the hippocampus repeatedly
47 reactivates these memories in concert with the neocortical representations (Grosmark &
48 Buzsáki, 2016; Ji & Wilson 2007; Khodagholy et al., 2017; Ólafsdóttir et al., 2016) thereby
49 strengthening and reorganising the representations in the neocortex (Marshall & Born, 2007;
50 McClelland et al., 1995). Reactivation of memory traces corresponds to sharp-wave/ripple
51 events evident in the hippocampal local field potential recordings during sleep (Diba &
52 Buzsáki, 2007) that coordinate with sleep spindles and sleep slow oscillations to drive active
53 systems consolidation (Clemens et al., 2007; Staresina et al., 2015; Khodagholy et al. 2017).
54 Although, sleep spindles and reactivation in form of sharp-wave/ripples can be enhanced to
55 compensate for large amounts of learning material (Gais et al., 2002; Mölle et al., 2009), it is
56 plausible that an active process of sleep on memory is limited by the amount of replay that
57 can be accommodated. This reasoning is supported by findings in humans that working
58 memory capacity predicts declarative memory retention across sleep (Fenn & Hambrick,
59 2012).
60   In accordance with that, Feld et al. (2016) recently showed that memory consolidation
61 of declarative content during sleep is limited in capacity. Here, participants learned either a
62 short (40), medium (160) or long (320) list of word-pairs. Participants in the medium
63 information load condition showed a large sleep-dependent memory advantage, whereas those
64 in the high information load condition no longer displayed a sleep benefit. This pattern of
65 results can be explained by a capacity limited process of active systems consolidation that
66 leads to local potentiation of memory traces that is accompanied by a more global process of
67 synaptic rescaling, that depotentiates synapses without being limited (Feld & Born, 2017;
68 Tononi & Cirelli, 2014). Extrapolating from this, at even higher information loads the limited
69 capacity for active systems consolidation is surpassed so that sleep would favour forgetting.
70   To test this, in the present study we doubled the information load from 320 to 640
71 word-pairs. We hypothesized, that sleep leads to more forgetting of word-pairs compared to a
72 wake group.

73 **2   Materials and Methods**

74 **3.1 Preregistration**

75 A rough outline of this study was preregistered at AsPredicted.org. It can be viewed through
76 this link http://aspredicted.org/blind.php?x=jc2y8t.

77 **3.2 Participants**

78 A total of 78 healthy, non-smoking, German-speaking participants performed the complete
79 study (two participants decided to drop out prematurely). They reported a regular wake-sleep
80 cycle, no intake of regular medication (except contraceptives) or illegal substances, and at
81 least the qualification to enter higher education. Beginning on the morning and throughout the
82 experiment the intake of caffeine- and alcohol-containing beverages was prohibited.
83 Participants were randomly assigned to either the wake condition ($N = 40$; 21 female, age

84   mean: 22.9 years, from 18-28 years) or sleep condition ($N = 38$; 19 female, age mean: 22.7
85   years, from 18-29 years). Participants received adequate monetary compensation for their
86   contribution and provided written informed consent prior to the experiment. The study was
87   approved by the local ethics committee.

3.3   Procedure

89   Participants arrived at 11:00 h and were seated in a room with four individual work-stations
90   that were positioned to minimize distractions from other participants. See Figure 1 for a
91   timeline of the experimental procedure. After a general instruction, participants completed a
92   working memory capacity test (automated operation span task, OSPAN; Unsworthet al.,
93   2005). From 12:00 to 17:00 h, participants learned the 640 word-pairs during a learning phase
94   divided into two parts of 320 word-pairs with a short snack break in between. The snack
95   consisted of a pretzel and one piece of fruit. Participants were asked not to actively rehearse
96   word-pairs during breaks and oral conversation was restricted. At 17:00 h, participants
97   received a standardized lunch consisting of either Pizza or Pasta. From 17:30 to 22:30 h, the
98   640 word-pairs were recalled (immediate recall) in two parts of 320 word-pairs each to
99   estimate, how many word-pairs had been successfully encoded. Again, participants received a
100   standardized snack in between the two parts. The snack consisted of two pieces of bread with
101   cheese or salami and a piece of fruit. At the beginning and at the end of each experimental day
102   as well as right before the immediate recall, the psychomotor vigilance task (PVT; Dinges et
103   al., 1997) as well as the Stanford Sleepiness Scale (SSS; Hoddes et al., 1973) were
104   administered (for details, see below). After recall, participants were assigned to either spend
105   the night in the laboratory watching standardized animal documentaries (wake group) or to
106   sleep at home (sleep group). Between 23:00 h on day 1 and 7:00 h on day 3 of the experiment,
107   all participants wore an actigraph (ActiLife v4.4.0, ActiGraph, Pensacola, FL, USA).
108   Participants staying in the laboratory were offered two snacks throughout the night and left
109   the laboratory at 7:00 h. These snacks consisted of a piece of fruit, a cereal bar and two slices
110   of raisin bread. All participants were instructed to refrain from sleeping during the day
111   following the first experimental day. After the wake group had a recovery night, the second
112   day of testing started at 8:00 h (delayed recall; around 33 hours after the end of the first
113   experimental day). All participants declared compliance to the sleep-wake schedule of the
114   experiment, which was ratified using actimetry data. At the end of the second experiment day
115   participants had to complete a word generation task (for details, see below).

3.4   Memory task

117   The word-pair task was implemented using Presentation® (version 16.3, Neurobehavioural
118   Systems, Berkeley, CA, USA) on computers running on Windows 7 and adapted from Feld et
119   al. (2016). Prior to each learning or recall block of 40 slightly related word-pairs, participants
120   were instructed how to perform the task followed by two mock trials of the task procedure.
121   Each phase consisted of 16 blocks which amounts to 640 word-pairs per phase. After 8 blocks
122   there was a longer, 30-minute break. During the learning phase, each word-pair was presented
123   once for 4s (1s inter-stimulus interval) on a horizontal axis divided by a hyphen. The left word
124   was always the cue word, whereas the right word was always the target word. The order in
125   which each word-pair occurred within a 40 word-pair block was randomized and the order of
126   the word-pair blocks was balanced between participants. The block order for each participant
127   was the same for the learning phase and both recall phases. Blocks lasted for 20 minutes. In
128   the learning phase this meant participants were actively encoding word-pairs for 200s (40
129   pairs x 5s), while the remaining 1000s were spent listening to relaxing audio files. During
130   recall, participants were presented the cue word and had 20s to type in their response. If

3

131   participants wanted to move on or could not remember the target word, they were able to skip
132   to the next word-pair by pressing the return key. Participants were instructed to answer even if
133   they were not certain of the answer, but to avoid guesses. The keyboard input was
134   immediately displayed which made it possible to correct for mistakes. Similar to the learning
135   phase, recall blocks were interweaved by periods during which participants listened to
136   relaxing audios. Again, blocks would start 20 minutes apart which resulted in a minimum of
137   400s of relaxing audio (if the participant used 20s to answer each cue word). The recall
138   procedure was identical for the immediate recall after learning and the delayed recall two
139   mornings later (Figure 1). Setting the duration of each block to 20 minutes provided two
140   benefits: It reduced interference between the blocks and ensured an equal amount of time
141   between each block during the learning phase and its' corresponding block during the
142   retrieval phase. Word-pairs were scored manually. If the answer contained spelling mistakes,
143   used the wrong gender or number, the answer was still checked as correct. In alignment with
144   Feld et al. (2016), we used absolute retention performance (the number of correctly recalled
145   words at immediate recall subtracted from the number of correctly recalled words at delayed
146   recall) as the dependent variable.

3.5   Working Memory (OSPAN)

148   The automated operation span (OSPAN; Unsworth et al., 2005) is a computer-based test to
149   assess working memory capacity. Participants are shown simple mathematical equations in
150   alternation with letters. Their task is to decide if the equations are correct while remembering
151   the letters in the order they were presented. After three to six trials an array of twelve letters is
152   shown and participants are instructed to click on the previously shown letters. For our analysis
153   we used the absolute score and a "partial load". The absolute score refers to the sum of all
154   correctly recalled letters, whereas the partial load is calculated by summing up only the letters
155   of the trials that were recalled in the correct order. One dataset was lost due to hardware
156   malfunction. This resulted in 38 datasets in the sleep condition and 39 datasets in the wake
157   condition.

3.6   Control measures (RWT, PVT, SSS, Actimetry)

159   In order to control for possible confounding variables, several measurements were taken. One
160   of the detrimental effects of sleep deprivation is a diminished verbal fluency due to impaired
161   retrieval processes (Harrison & Horne, 1997). To investigate, if, even after a recovery night,
162   the wake condition was still suffering the negative consequences of sleep deprivation,
163   participants completed a word generation task (RWT: Regensburger Wortflüssigkeits-Test)
164   (Aschenbrenner et al., 2000). The assignment was to generate as many words as possible of a
165   given category (hobbies), or to generate words that start with a specific letter (in this case
166   "m") during a 2-minute period. For our analysis, we added up both results for a combined
167   score. The PVT measures the participant's mean reaction speed and is an indicator of
168   vigilance. The 5-minute test requires pressing the space bar, as soon as a bright red
169   millisecond timer appears on the screen of a computer and starts counting up from 0000 in
170   milliseconds immediately. The subject's reaction time is displayed as soon as the space bar is
171   pressed. For our analysis, we calculated the mean reaction speed (1 / reaction time) for each
172   participant and the number of lapses, defined as reaction speed above 500ms (Dinges &
173   Kribbs, 1991). Due to hard drive malfunction, we lost 6 datasets (3 in each condition) for the
174   psychomotor vigilance task (PVT). The SSS consists of a single item measuring the
175   subjective sleepiness on a scale of 1 ("Feeling active, vital, alert or wide awake") to 8
176   ("Asleep") (Hoddes et al., 1973). Five actimetry datasets (one from the sleep condition) could
177   not be recovered due to hardware malfunction.

4

3.7 Statistical analysis

All statistical analysis was conducted using SPSS Version 22.0.0 on a computer running on Windows 7 and Jamovi Version 1.0.4.0. Unless stated otherwise, we relied on a univariate analysis of variance (ANOVA) with experimental condition and sex as independent variables. The SSS and the PVT were analysed using a repeated measures ANOVA with each of the five data points as a within-subjects factor and condition and sex as between-subject factors. The significance threshold for all statistical tests was set at 0.05.

3 Results

3.1 Memory Performance

Recall in the wake group decreased from $M_{Wake} = 117.85$ ($SEM_{Wake} = 7.16$) at immediate recall to $M_{Wake} = 106.65$ ($SEM_{Wake} = 6.78$) at delayed recall. In the sleep group the number of remembered word-pairs decreased from $M_{Sleep} = 112.08$ ($SEM_{Sleep} = 9.93$) to $M_{Sleep} = 103.61$ ($SEM_{Sleep} = 9.33$) at delayed recall. In a confirmatory analysis the sleep and the wake group did not differ in retention performance (the absolute amount of word-pairs that were forgotten from the immediate recall phase to the delayed recall) on the word-pair task ($M_{Wake} = -11.2$ $SEM_{Wake} = 1.73$, $M_{Sleep} = -8.5$ $SEM_{Sleep} = 1.70$; sleep/wake: $F_{1,74} = 1.25$, $p = .27$; sex: $F_{1,74} = 0.012$, $p = 0.91$; Figure 2). To provide an estimate of the evidence for the null effect we performed an equivalence test using the two one sided test procedure (Lakens et al., 2018) with upper and lower effect size bound set at $d = \pm0.2$. Statistically this means we tested whether the hypotheses that the positive effect of sleep on memory retention is larger than $d = 0.2$ and that the negative effect of sleep on memory retention is larger than $d = 0.2$ can be rejected with a one sided t-tests each. Here, there was evidence that a detrimental effect of sleep on retention larger than $d = -0.20$ can be ruled out ($t_{76} = 2.01$, $p = 0.02$), whereas, there was no evidence against a positive effect of sleep larger than $d = 0.20$ ($t_{76} = 0.24$, $p = 0.60$). This means that for sleep induced forgetting there is evidence against medium and large effects, whereas small effects ($d \leq 0.2$) could exist in our paradigm. For sleep-dependent consolidation even effects larger than $d = 0.2$ cannot be ruled in out. In addition to this analysis, following the null-hypothesis-testing (NHST) tradition, we also used Bayesian statistics to determine evidence for null effects. This analysis provides the likelihood of the model given the data, rather than the probability of the data given the model as in NHST. Similar to the equivalence test, calculating the one-sided Bayes factor provided evidence against a detrimental effect of sleep on memory, it came up with $BF_{01} = 8.26$ in favour of the null, whereas, evidence against a positive effect of sleep versus the null was undecided $BF_{01} = 1.45$. This means that the null hypothesis of no detrimental effect of sleep on memory is 8.26 times more likely than the alternative hypothesis of a detrimental effect of sleep on memory in comparison with our wake condition.

Next, we investigated whether these results were due to either group having learned more words during the initial learning phase. Although the wake group descriptively learned slightly more word-pairs, this difference did not reach statistical significance (sleep/wake: $F_{1,74} = 0.17$, $p = .68$).

3.2 Memory Performance Blockwise

To verify this null-result, we next explored the performance per block for each condition (the experiment consisted of 16 blocks of 40 word-pairs each). As before, the absolute difference was used as a dependent variable in a repeated measures ANOVA with the block (1 to 16) as the repeated measurement and sleep/wake condition and sex as between subject variables (see

223 Figure 3). However, no significant result emerged from this analysis (sleep/wake: $F_{1,71} = 1.47$,
224 $p = .22$; sex: $F_{1,71} = 0.083$, $p = .77$; sleep/wake * block: $F_{15,57} = 0.99$, $p = 0.48$; sex * block:
225 $F_{15,57} = 0.83$, $p = 0.64$).

3.3 Gains and Losses

227 Gains refer to words that were correctly recalled during delayed recall, but not during
228 immediate recall. Losses accordingly refer to words that were correctly remembered during
229 immediate recall, but not during delayed recall. Subjects in the sleep and wake condition both
230 gained on average 12.3 words ($SEM_{Wake} = 0.90$; $SEM_{Sleep} = 0.87$). Participants in the wake
231 group lost on average 23.5 words ($SEM_{Wake} = 1.49$), while participants in the sleep group lost
232 on average 20.79 words ($SEM_{Sleep} = 1.62$). We found no statistically significant difference
233 between groups for neither gains (sleep/wake: $F_{1,74} = 0.001$, $p = .97$), nor losses (sleep/wake:
234 $F_{1,74} = 1.49$, $p = .23$; Table 1).

3.4 Wrong Answers

236 Next, we analysed the incorrect answers in detail. We investigated if there were group
237 differences between the number of repeated wrong answers (same wrong answer during
238 encoding and retrieval). This was not the case ($M_{Sleep} = 114.03$ ($SEM_{Sleep} = 8.64$), $M_{Wake} =$
239 105.85 ($SEM_{Wake} = 6.36$), sleep/wake: $F_{1,74} = 0.64$, $p = .43$; Table 1). Additionally, we
240 explored whether the occurrence of empty responses differed between sleep/wake groups. We
241 found no statistically significant difference between groups when considering the difference
242 in absolute number of empty responses between immediate and delayed recall ($M_{Sleep} = -7.42$
243 ($SEM_{Sleep} = 10.32$), $M_{Wake} = 7.83$ ($SEM_{Wake} = 14.70$), sleep/wake: $F_{1,74} = 0.84$, $p = .36$; Table
244 1) or when considering the proportion of empty responses in all wrong answers ($M_{Sleep} = -$
245 0.024 ($SEM_{Sleep} = 0.018$), $M_{Wake} = 0.001$ ($SEM_{Wake} = 0.026$), sleep/wake: $F_{1,74} = 0.78$, $p = .38$;
246 Table 1). Likewise, there was no significant difference between groups regarding the
247 difference in number of incorrect responses at immediate and delayed recall. (excluding
248 empty responses) ($M_{Sleep} = 15.89$ ($SEM_{Sleep} = 9.59$), $M_{Wake} = 3.38$ ($SEM_{Wake} = 14.10$),
249 sleep/wake: $F_{1,74} = 0.65$, $p = .42$; Table 1).
250   Word-pool errors refer to wrong answers containing a word that was learned at
251 another point in the experiment (either as a cue word or as a target word). Using the relative
252 occurrence within all wrong answers, we calculated the difference between word-pool errors
253 during delayed recall and during immediate recall. We found no statistically significant
254 difference between groups ($M_{Sleep} = 0.012$ ($SEM_{Sleep} = 0.009$), $M_{Wake} = -0.006$ ($SEM_{Wake} =$
255 0.009), sleep/wake: $F_{1,74} = 1.83$, $p = .18$; Table 1).

3.5 Working Memory Capacity Task

257 We used a MANOVA because the two dependent variables correlated highly ($r_{75} = 0.908$, $p <$
258 .01). The MANOVA revealed no difference in the OSPAN$_{Absolute}$ score and the OSPAN$_{Partial}$
259 score between sleep and wake condition ($M_{Absolute, Sleep} = 43.18$, $SEM_{Absolute, Sleep} = 2.73$;
260 $M_{Absolute, Wake} = 46.64$, $SEM_{Absolut, Wake} = 2.88$; $M_{Partial, Sleep} = 60.39$, $SEM_{Partial, Sleep} = 1.57$;
261 $M_{Partial, Wake} = 61.26$, $SEM_{Partial, Wake} = 1.99$; sleep/wake: OSPAN$_{Absolute}$: $F_{1,73} = 0.84$, $p = .36$;
262 OSPAN$_{Partial}$: $F_{1,73} = 0.12$, $p = .73$; Table 2). To investigate the relation between working
263 memory capacity and memory performance on the word-lists (using the difference score), we
264 calculated a Pearson product-moment correlation between memory retention performance and
265 working memory capacity separately for the sleep and wake conditions, and for the OSPAN
266 absolute and partial score. There was a trend towards significance when looking at the wake
267 condition and the absolute OSPAN score ($r_{39} = -.31$, $p = .059$), while all other correlations

268 were insignificant ($p > .11$). However, when considering all participants in both groups, there
269 was a statistically relevant negative relation between retention performance and the absolute
270 OSPAN score ($r_{75} = -.27$, $p = .02$).

3.6 Control Measures (Actimetry, SSS, PVT, RWT)

3.6.1 Actimetry Data

273 To analyse actimetry data, we used a repeated measures ANOVA to compare both groups'
274 activity levels in four 8h windows. We found a statistically significant sleep/wake * time
275 interaction ($F_{1.92, 136,55} = 35.1$, $p < 0.001$, $\eta^2 = .33$; Greenhouse-Geisser corrected). Subsequent
276 two-tailed t-tests revealed, that the difference was mainly driven by a difference in the first 8h
277 window, i.e. during sleep deprivation in the wake group ($t_{48.0} = -21.2$, $p < .001$; Figure 1b).

3.6.2 Subjective Sleepiness (SSS)

279 The SSS score at some time points was affected by the sleep/wake condition (sleep/wake *
280 time: $F_{4,304} = 2.46$, $p = .046$, $\eta^2 = .031$). Individual two-tailed t-tests revealed that this
281 difference was mainly driven by higher subjective sleepiness in the wake group during the
282 fourth ($t_{76} = -1.92$, $p = .058$) and fifth ($t_{76} = -1.97$, $p = .053$) assessment points, which both
283 occurred on the second day of the experiment (Table 2).

3.6.3 Vigilance (PVT)

285 There was no significant difference between groups' average reaction speed (all $p > .24$).
286 Likewise, there was no significant main effect or interaction of condition regarding the
287 number of lapses (all $p > .44$; Table 2).

3.6.4 Word generation task (RWT)

289 Analyses revealed no significant difference between conditions in the word generation task
290 ($F_{1,74} = 0.54$, $p = .47$; Table 2).

4 Discussion

292 Previous work suggests that sleep-dependent memory consolidation is a process limited in
293 capacity and that learning large amounts of information overloads active systems
294 consolidation and abolishes the positive effect of sleep on memory retention (Feld et al.
295 2016). Extrapolating from this data, we hypothesized that at even higher loads of information
296 during encoding sleep may favour forgetting over consolidation (Feld & Born 2017). Here,
297 we directly tested this hypothesis by asking participants to encode a very large amount of
298 information (640 word-pairs, twice the amount used before in the long list condition of Feld et
299 al., 2016) before either a night of sleep or total sleep deprivation. Contrary to our predictions,
300 we found word-pair retention to be comparable between the sleep and wake groups. While
301 this null-effect can be reconciled with the view that capacities for consolidating memory
302 during sleep are limited, it contradicts our hypothesis that sleep causes increased forgetting of
303 declarative memory, when compared to wakefulness. It is important to note that sleep might
304 still induce forgetting under conditions of massed learning, but that this effect might be
305 masked by a direct comparison with a wake condition which itself induces forgetting (as
306 discussed in detail below). A thorough post-hoc analysis revealed no group differences
307 regarding a multitude of response patterns (such as gain and loss, word-pool errors and wrong
308 answers).

309   In light of these results we propose that forgetting of memory traces could be achieved
310   by different processes for the wake and sleep group, respectively. In the wake group, memory
311   traces are more prone to interference, whereas during sleep memories are protected from
312   interference (Ellenbogen, Hulbert et al., 2006, Wixted, 2005). However, in the sleep group,
313   memories might be forgotten due to global synaptic downscaling (Tononi & Cirelli, 2006). In
314   this scenario, our failure to find sleep enhanced forgetting would be explained by wake
315   forgetting accelerating at a similar pace making it very difficult to dissociate these processes.
316   Accordingly, if forgetting in the wake condition is primarily driven by interference within the
317   task, then increasing or decreasing this interference (e.g. by making the stimulus material
318   more or less semantically related) will lead to more forgetting in participants in the wake
319   condition, but not in the sleep condition (Drosopoulos et al., 2007; Yonelinas et al., 2019).
320   Likewise, task-unrelated interference could be manipulated by asking participants in the wake
321   condition to learn an unrelated verbal memory task during sleep deprivation. Contrasting with
322   the wake state, we assume that global synaptic downscaling causes forgetting in the sleep
323   group, which could equally be manipulated in this paradigm. Since global synaptic
324   downscaling is assumed to occur primarily during slow waves (Kim et al., 2019; Vyazovskiy
325   et al., 2008), closed loop auditory stimulation could be employed to increase slow waves,
326   causing more forgetting in participants that have previously encoded a high amount of
327   information (Ngo et al., 2013). Conversely, preventing participants from reaching deep sleep
328   should lead to less forgetting. This would prevent forgetting in two ways, first by preventing
329   interference through new encoding, as long as sleep itself is maintained, and second by
330   preventing synaptic downscaling during deep sleep.
331   An alternative account of the absence of an enhanced forgetting during sleep, induced
332   by massed learning, can be derived by considering recent findings of synaptic downscaling
333   mechanisms during sleep. Although it has been suggested that active systems consolidation is
334   specific and selective, whereas synaptic downscaling is global and indiscriminate (e.g. Feld &
335   Born, 2017), there also exists the opposite suggestion of selective downscaling during sleep
336   (Tononi & Cirelli 2014). This latter account is largely based on findings of selective
337   weakening of synapses during sleep, where only weaker, more plastic, synapses are erased,
338   while stronger synapses remain stable (De Vivo et al., 2017). In addition, it has been found
339   that sharp-wave/ripples (a correlate of memory reactivation during sleep) are involved in the
340   depotentiation of synapses within the hippocampus (Norimoto et al., 2018), whereas they
341   appear to be involved in the potentiation of synapses in the cortex during sleep spindles
342   (Khodagholy et al., 2017). This offers the intriguing possibility that active systems
343   consolidation and selective synaptic downscaling during sleep occur in a highly coordinated
344   fashion, i.e., during the same reactivation events but in different brain areas. According to this
345   framework the successful integration of memories into the knowledge stores of the cortex via
346   active systems consolidation would signal the deletion of redundant memory traces from the
347   hippocampus through synaptic downscaling. In conclusion, similar to active systems
348   consolidation, selective synaptic downscaling during sleep might be limited by the number of
349   available reactivations during sleep. This offers an explanation for the lack of sleep-induced
350   net forgetting.
351   Turning to our findings on working memory, prior work by Fenn and Hambrick
352   (2012) reported a positive relation between working memory and sleep related memory
353   performance. In contrast to that, Feld et al. (2016) measured working memory performance
354   before any sleep/wake intervention took place (which rules out any biases due to the circadian
355   rhythms or sleep manipulation) and found no significant correlation between the two. Using
356   the same methodology, in the present study, we found a negative correlation, although most
357   did not reach significance. Given these contradicting findings, we suggest that sleep-
358   dependent memory performance is likely independent of working memory functioning.

8

359   There are several limitations that were impossible to eliminate in this study and
360 therefore possibly contribute to our results. (1) Variability in memory performance between
361 subjects increases with list size and although the sample size was large in comparison to other
362 studies (Ellenbogen, Hulbert et al., 2006; Igloi et al., 2015; Marshall et al., 2006; Ngo et al.,
363 2013; Studte et al., 2017; Quigley et al., 2000) this probably decreased statistical power. To
364 ameliorate this noise issue, criterion learning could be used in future studies, where learning is
365 repeated until a certain percentage of word-pairs can be recalled correctly. Importantly, a
366 study comparing different criterions found that a 60% criterion is well suited to tap into the
367 sleep effect (Drosopoulos et al., 2007). We did not use this method, as it would have
368 consumed significantly more time for learning, which would have made it impossible to space
369 out learning und thus strongly increase interference effects. (2) It is possible, that not enough
370 time had passed for sleep effects to emerge. Already Graves (1936) using nonsense-syllables
371 found a sleep benefit only after 72h and not at shorter intervals. Similarly, Richardson and
372 Gough (1963) did not find a sleep effect after 24h/48h, but after 144h. Especially, for large
373 amounts of information it is conceivable that consolidation as well as forgetting is carried
374 over to subsequent nights. (3) We tested declarative memories that were intentionally
375 encoded. It might be that the underlying processes, such as an enhanced activation of
376 prefrontal-hippocampal circuitry, preclude such information from sleep-dependent forgetting
377 (Himmer et al., 2017), which stimulates the idea to compare, in future studies, sleep effects on
378 high loads of intentionally and incidentally encoded memory.
379   To conclude, in the current experiment we did not find evidence that a high
380 information load leads to more forgetting during sleep when compared to wakefulness. These
381 findings can be explained by different mechanisms leading to forgetting in both brain states:
382 interference-induced forgetting in the wake group and forgetting due to global synaptic
383 downscaling in the sleep group. We propose several approaches how future studies can test
384 this new hypothesis.

Acknowledgments

Declaration of conflicting interests

391 The authors declared that they have no conflicts of interest with respect to their authorship or
392 the publication of this article.

References

394 Abraham, W. C., & Robins, A. (2005). Memory retention–the synaptic stability versus
395   plasticity dilemma. *Trends in Neurosciences, 28*(2), 73-78.
396 Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. (2018). Raincloud plots: a
397   multi-platform tool for robust data visualization. *PeerJ Preprints, 6*, e27137v1.
398 Aschenbrenner, S., Tucha, O., & Lange, K. W. (2000). *RWT: Regensburger Wortflüssigkeits-
399   Test*. Göttingen: Hogrefe.
400 Battaglia, F. P., Benchenane, K., Sirota, A., Pennartz, C. M., & Wiener, S. I. (2011). The
401   hippocampus: hub of brain network communication for memory. *Trends in Cognitive
402   Sciences, 15*(7), 310-318.

Clemens, Z., Mölle, M., Erőss, L., Barsi, P., Halász, P., & Born, J. (2007). Temporal coupling of parahippocampal ripples, sleep spindles and slow oscillations in humans. *Brain, 130*(11), 2868-2878.

De Vivo, L., Bellesi, M., Marshall, W., Bushong, E. A., Ellisman, M. H., Tononi, G., & Cirelli, C. (2017). Ultrastructural evidence for synaptic scaling across the wake/sleep cycle. *Science, 355*(6324), 507-510.

Diba, K., & Buzsáki, G. (2007). Forward and reverse hippocampal place-cell sequences during ripples. *Nature Neuroscience, 10*(10), 1241-1242

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience, 11*, 114–126. doi:10.1038/nrn2762

Dinges D., & Kribbs N. B. (1991). Performing while sleepy: effects of experimentally-induced sleepiness. In: T. H. Monk (Ed.), *Human performance and cognition. Sleep, sleepiness and performance* (pp. 97-128). Oxford, England: John Wiley.

Dinges, D. F., Pack, F., Williams, K., Gillen, K. A., Powell, J. W., Ott, G. E., . . . Pack, A. I. (1997). Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 hours per night. *Sleep, 20*(4), 267-277. doi:10.1093/sleep/20.4.267

Drosopoulos, S., Schulze, C., Fischer, S., & Born, J. (2007). Sleeps function in the spontaneous recovery and consolidation of memories. *Journal of Experimental Psychology: General, 136*(2), 169-183. doi:10.1037/0096-3445.136.2.169

Ellenbogen, J. M. Hulbert, J. C., Stickgold, R., Dinges, D. F., & Thompson-Schill, S. L. (2006). Interfering with theories of sleep and memory: Sleep, declarative memory, and associative interference. *Current Biology, 16*, 1290-1294.

Ellenbogen, J. M., Payne, J. D., & Stickgold, R. (2006). The role of sleep in declarative memory consolidation: passive, permissive, active or none?. *Current Opinion in Neurobiology, 16*(6), 716-722.

Feld, G. B., & Born, J. (2017). Sculpting memory during sleep: concurrent consolidation and forgetting. *Current Opinion in Neurobiology, 44*, 20-27. doi:10.1016/j.conb.2017.02.012

Feld, G. B., Weis, P. P., & Born, J. (2016). The limited capacity of sleep-dependent memory consolidation. *Frontiers in Psychology, 7*, 1-12. doi:10.3389/fpsyg.2016.01368

Fenn, K. M., & Hambrick, D. Z. (2012). Individual differences in working memory capacity predict sleep-dependent memory consolidation. *Journal of Experimental Psychology: General, 141*(3), 404-410. doi:10.1037/a0025268

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

Graves, E. A. (1936). The effect of sleep on retention. *Journal of Experimental Psychology, 19*(3), 316-322.

Grosmark, A. D., & Buzsáki, G. (2016). Diversity in neural firing dynamics supports both rigid and learned hippocampal sequences. *Science, 351*(6280), 1440-1443.

Harrison, Y., & Horne, J. A. (1997). Sleep deprivation affects speech. *Sleep, 20*(10), 871-877.

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

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of sleepiness: *A new approach. Psychophysiology, 10*(4), 431-436. doi:10.1111/j.1469-8986.1973.tb00801.x

Igloi, K., Gaggioni, G., Sterpenich, V., & Schwartz, S. (2015). A nap to recap or how reward regulates hippocampal-prefrontal memory networks during daytime sleep in humans. *Elife, 4*, e07903.

Jenkins, J. G., & Dallenbach, K. M. (1924). Obliviscence during sleep and waking. *The American Journal of Psychology, 35*(4), 605-612.

Ji, D., & Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience, 10*(1), 100.

Khodagholy, D., Gelinas, J. N., & Buzsáki, G. (2017). Learning-enhanced coupling between ripple oscillations in association cortices and hippocampus. *Science, 358*(6361), 369-372.

Kim, J., Gulati, T., & Ganguly, K. (2019). Competing roles of slow oscillations and delta waves in memory consolidation versus forgetting. *Cell, 179*(2), 514-526.

Klinzing, J. G., Niethard, N., & Born, J. (2019). Mechanisms of systems memory consolidation during sleep. *Nature Neuroscience, 22*(10), 1598-1610.

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science, 1*(2), 259-269.

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society London B, 262,* 23-81. doi:10.1098/rstb.1971.0078

Marshall, L., & Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends in Cognitive Sciences, 11*(10), 442-450. doi:10.1016/j.tics.2007.09.001

Marshall, L., Helgadóttir, H., Mölle, M., & Born, J. (2006). Boosting slow oscillations during sleep potentiates memory. *Nature, 444*(7119), 610.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*(3), 419-457. doi:10.1037/0033-295x.102.3.419

Mölle, M., Eschenko, O., Gais, S., Sara, S. J., & Born, J. (2009). The influence of learning on sleep slow oscillations and associated spindles and ripples in humans and rats. *European Journal of Neuroscience, 29*(5), 1071-1081.

Ngo, H. V. V., Martinetz, T., Born, J., & Mölle, M. (2013). Auditory closed-loop stimulation of the sleep slow oscillation enhances memory. *Neuron, 78*(3), 545-553.

Norimoto, H., Makino, K., Gao, M., Shikano, Y., Okamoto, K., Ishikawa, T., ... & Ikegaya, Y. (2018). Hippocampal ripples down-regulate synapses. *Science, 359*(6383), 1524-1527.

Ólafsdóttir, H. F., Carpenter, F., & Barry, C. (2016). Coordinated grid and place cell replay during rest. *Nature Neuroscience, 19*(6), 792.

Quigley, N., Green, J. F., Morgan, D., Idzikowski, C., & King, D. J. (2000). The effect of sleep deprivation on memory and psychomotor function in healthy volunteers. *Human Psychopharmacology: Clinical and Experimental, 15*(3), 171-177. doi:10.1002/(sici)1099-1077(200004)15:3<171::aid-hup155>3.0.co;2-d

Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Reviews, 93*(2), 681-766. doi:10.1152/physrev.00032.2012

Richardson, A., & Gough, J. E. (1963). The long range effect of sleep on retention. *Australian Journal of Psychology, 15*(1), 37-41.

Robins, A. (1995). Catastrophic forgetting, rehearsal and pseudorehearsal. *Connection Science, 7*(2), 123-146.

Staresina, B. P., Bergmann, T. O., Bonnefond, M., Van Der Meij, R., Jensen, O., Deuker, L., ... & Fell, J. (2015). Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. *Nature Neuroscience, 18*(11), 1679.

Studte, S., Bridger, E., & Mecklinger, A. (2017). Sleep spindles during a nap correlate with post sleep memory performance for highly rewarded word-pairs. *Brain and Language, 167*, 28-35.

11

501 Tononi, G., & Cirelli, C. (2006). Sleep function and synaptic homeostasis. *Sleep Medicine*
502   *Reviews, 10*(1), 49-62. doi:10.1016/j.smrv.2005.05.002
503 Tononi, G., & Cirelli, C. (2014). Sleep and the price of plasticity: from synaptic and cellular
504   homeostasis to memory consolidation and integration. *Neuron, 81*(1), 12-34.
505 Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the
506   operation span task. *Behavior Research Methods, 37*(3), 498-505. doi:10.3758/bf03192720
507 Vyazovskiy, V. V., Cirelli, C., Pfister-Genskow, M., Faraguna, U., & Tononi, G. (2008).
508   Molecular and electrophysiological evidence for net synaptic potentiation in wake and
509   depression in sleep. *Nature Neuroscience, 11*(2), 200-208.
510 Walker, M. P., & Stickgold, R. (2010). Overnight alchemy: sleep-dependent memory evolution.
511   *Nature Reviews Neuroscience, 11*(3), 218.
512 Wixted, J. T. (2005). A theory about why we forget what we once knew. *Current Directions in*
513   *Psychological Science, 14*(1), 6-9.
514 Yonelinas, A. P., Ranganath, C., Ekstrom, A. D., & Wiltgen, B. J. (2019). A contextual binding
515   theory of episodic memory: systems consolidation reconsidered. *Nature Reviews*
516   *Neuroscience, 20*(6), 364-375.
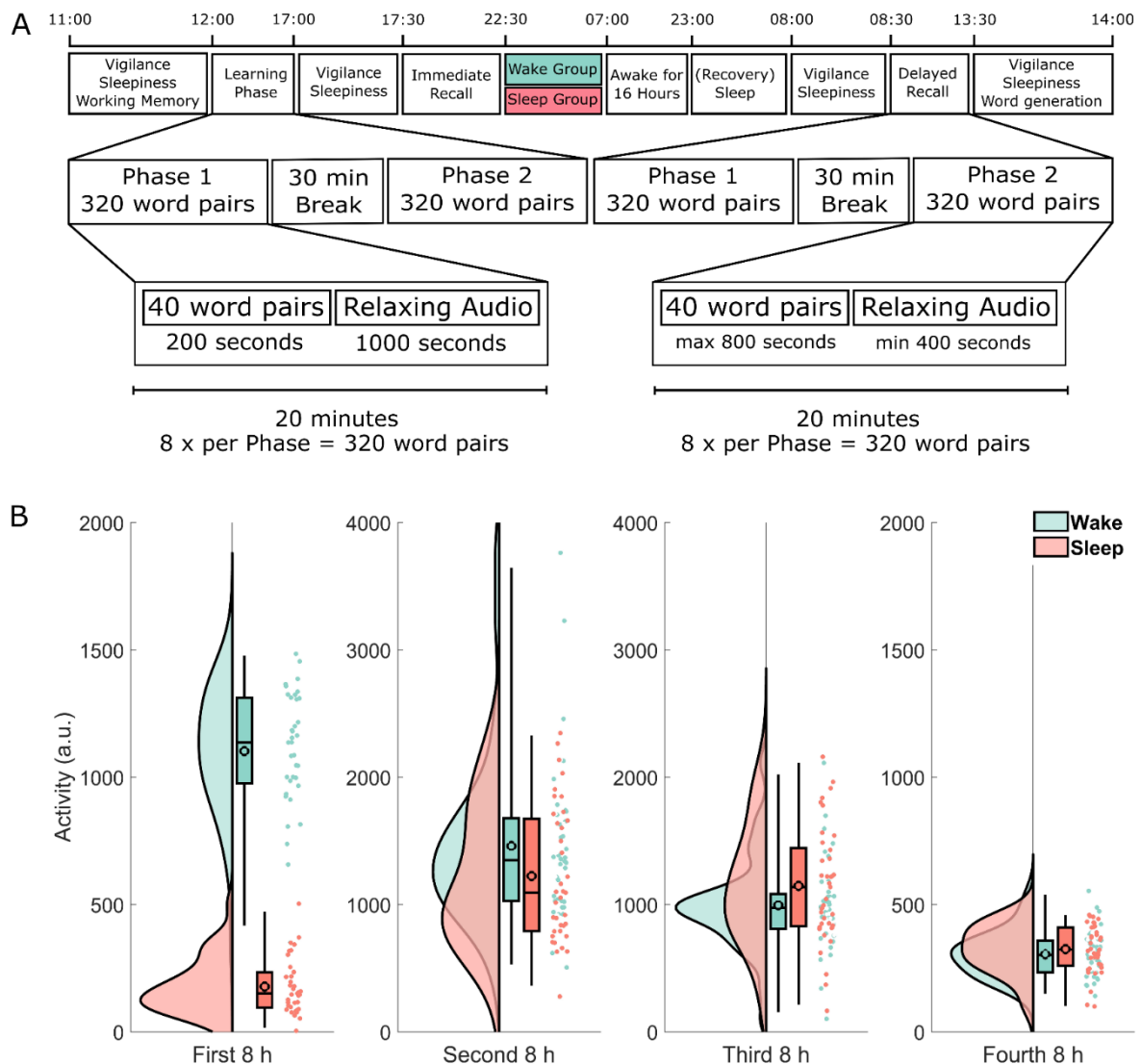
Figure 1. **(A)** Timeline of the experimental procedure. Learning started at 12:00 h and was followed by an immediate recall at 17:30 h as well as a delayed recall at 8:00 h two days later. The learning and recall phases each took five hours and consisted of two parts each roughly 2 hours 15 minutes long, separated by a 30 minute break. During each part, participants either learned or retrieved 320 word-pairs for a total of 640 word-pairs. Each part was further divided into eight blocks of 40 word-pairs. Each block took exactly 20 minutes. During the learning phase participants spent 3 minutes and 20 seconds per block learning word-pairs one at a time and then listened to 16 minutes and 40 seconds of relaxing audio files. During the recall phase participants had up to 20 seconds (a maximum of 13 minutes and 20 seconds per block) to respond to each of the sequentially shown cue words by typing the correct target word. Participants spent the remaining time listening to a relaxing audio file. Note that both recall phases (immediate and delayed recall) followed exactly the same procedure. **(B)** Actimetry data. Each participant was given an actigraph at the end of immediate recall to verify compliance. The y-axis shows the activity of each participant during each of the four eight-hour periods in arbitrary units. Each raincloud plot consists of the estimated distribution, a box-plot (indicating the median and the 2%, 25%, 98% quantiles, the black outlined circles depict the mean) and the activity estimations for each subject as an individual point. Data from participants in the wake group is shown in green, whereas data from the sleep group is shown

13

535 in red. Note the different scale used here. See Allen et al. (2018) for the code used in this
536 visualisation.

537

538
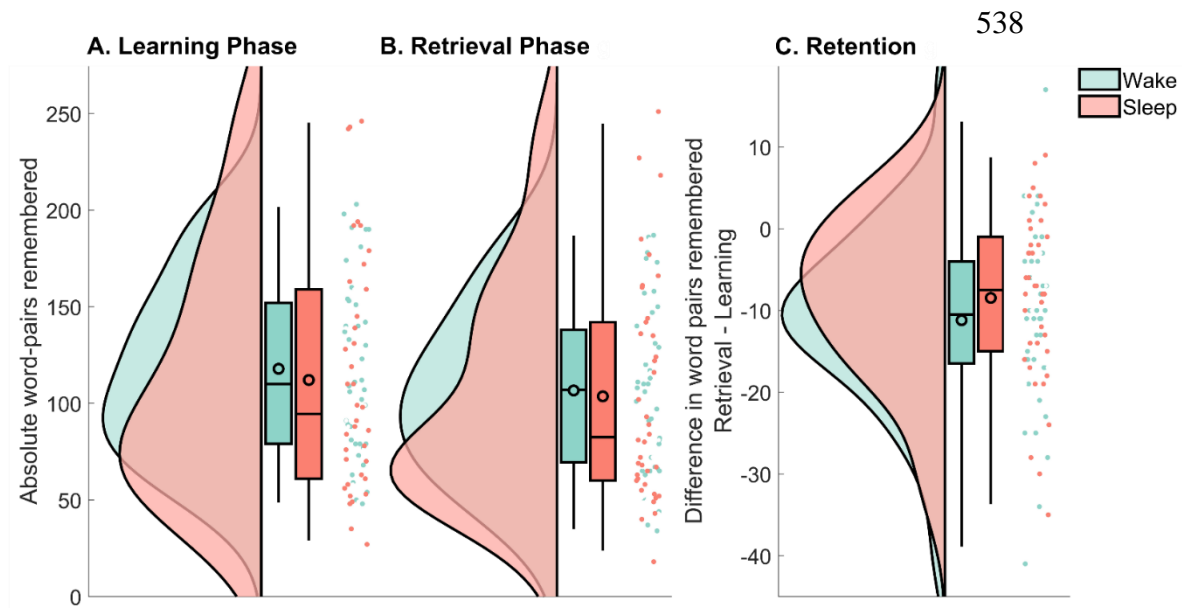
539 Figure 2. Raincloud plots (curves depict the estimated distribution, box-plots provide the
540 median and the 2%, 25%, 75% and 98% quantiles, the black outlined circles depict the mean,
541 the dots show individual data points) of the remembered word-pairs for the wake group (green)
542 and the sleep group (red). **(A)** The absolute amount recalled during immediate recall that was
543 part of the learning phase. **(B)** The absolute amount recalled during delayed recall that was part
544 of the retrieval phase. **(C)** The difference in word-pairs remembered between retrieval and
545 learning. Note the different scale used here. See Allen et al. (2018) for the code used in this
546 visualisation.

547

548
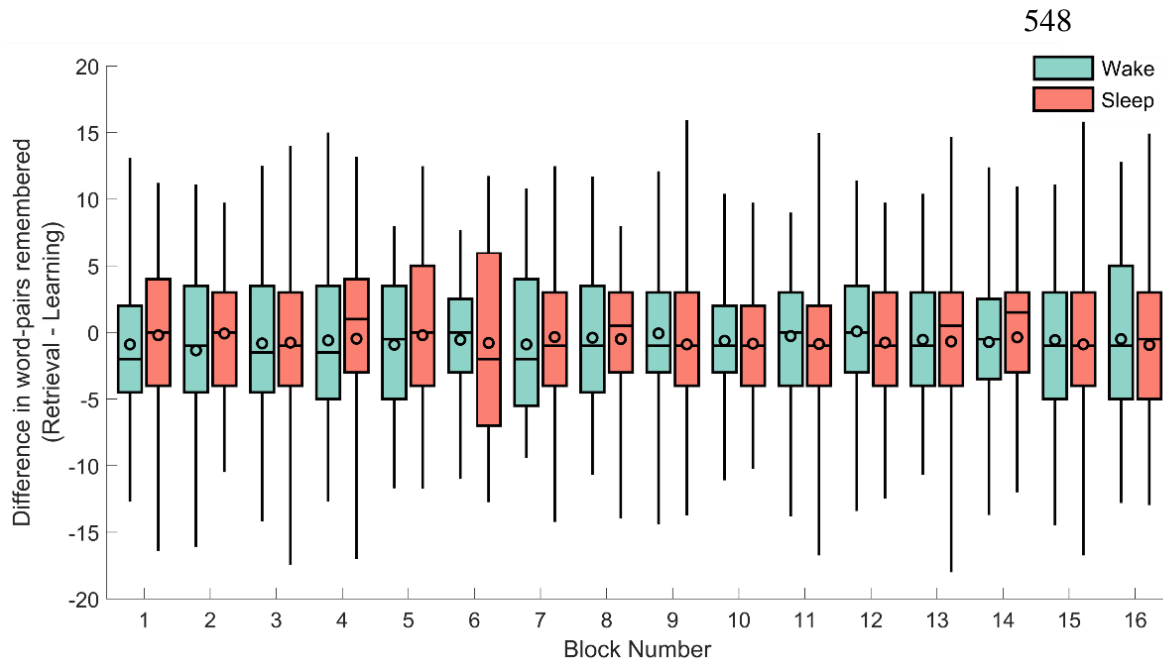
549 Figure 3. The difference in remembered word-pairs between the retrieval and learning phase
550 for each of the individual 16 blocks for the sleep (red) and wake groups (green). The horizontal
551 line of each box-plot indicates the median, the black outlined circles depict the mean, the border
552 of the box indicates the 25% and 75% quartiles and the whiskers the 2% and 98% quantiles
553 respectively.

554 Table 1. Mean and SEM (in brackets) for several measures of patterns within wrong answers.
555 Contrasted between the wake group and the sleep group.

	Wake Group Mean (SEM)	Sleep Group Mean (SEM)
Gains[1]	12.3 (0.90)	12.3 (0.87)
Loss[2]	23.5 (1.49)	20.79 (1.62)
Number of repeated wrong answers	105.85 (6.36)	114.03 (8.64)
Number of empty responses[3]	7.83 (14.70)	-7.42 (10.32)
Proportion of empty responses[3]	0.001 (0.026)	-0.024 (0.018)
Incorrect responses[3] (excluding empty responses)	3.38 (14.10)	15.89 (9.59)
Word-pool error[3, 4]	0.012 (0.009)	-0.006 (0.009)

556 [1] Correctly recalled during delayed recall, but not during immediate recall
557 [2] Correctly recalled during immediate recall, but not during delayed recall
558 [3] Difference between immediate and delayed recall
559 [4] Wrong answer was learned at another point in the experiment

560 Table 2. Mean and SEM (in brackets) for each of the control measures and the working
561 memory test for the wake group and the sleep group.

	Wake Group Mean (SEM)	Sleep Group Mean (SEM)
SSS – 1[1]	2.25 (0.14)	2.32 (0.14)
SSS – 2[1]	3.15 (0.15)	3.42 (0.14)
SSS – 3[1]	4.43 (0.19)	4.32 (0.17)
SSS – 4[1]	3.28 (0.17)	2.84 (0.15)
SSS – 5[1]	3.42 (0.20)	2.89 (0.18)
PVT – 1[2]	3.19 (0.033)	3.29 (0.038)
PVT – 2[2]	3.22 (0.038)	3.29 (0.041)
PVT – 3[2]	3.21 (0.048)	3.24 (0.046)
PVT – 4[2]	3.23 (0.038)	3.26 (0.051)
PVT – 5[2]	3.25 (0.047)	3.33 (0.044)
PVT – 1 (lapses)	0.55 (0.15)	0.68 (0.12)
PVT – 2 (lapses)	0.53 (0.12)	0.58 (0.13)
PVT – 3 (lapses)	0.63 (0.28)	0.74 (0.22)
PVT – 4 (lapses)	0.77 (0.30)	0.77 (0.20)
PVT – 5 (lapses)	0.74 (0.30)	0.46 (0.17)
OSPAN (absolute)	46.6 (2.88)	43.2 (2.7)
OSPAN (partial)	61.3 (1.99)	60.4 (1.6)
RWT (total score)	41.3 (1.14)	39.9 (1.3)

562

[1] Stanford Sleepiness Scale from 1 (wide awake) to 8 (asleep)
[2] Reaction speed (defined as 1 / reaction time)

17