Behavioral/Cognitive

The Role of Sleep in Motor Sequence Consolidation: Stabilization Rather Than Enhancement

Almut Nettersheim,[1] **Manfred Hallschmid,**[2,3,4] **Jan Born,**[2,3,4] **and Susanne Diekelmann**[2]

[1]Department of Neuroendocrinology, University of Lübeck, D-23538 Lübeck, Germany, [2]Department of Medical Psychology and Behavioral Neurobiology, University of Tübingen, D-72076 Tübingen, Germany, [3]German Center for Diabetes Research, D-72076 Tübingen, Germany, and [4]Institute for Diabetes Research and Metabolic Diseases of the Helmholtz Center Munich at the University of Tübingen, D-72076 Tübingen, Germany

Sleep supports the consolidation of motor sequence memories, yet it remains unclear whether sleep stabilizes or actually enhances motor sequence performance. Here we assessed the time course of motor memory consolidation in humans, taking early boosts in performance into account and varying the time between training and sleep. Two groups of subjects, each participating in a short wake condition and a longer sleep condition, were trained on the sequential finger-tapping task in the evening and were tested (1) after wake intervals of either 30 min or 4 h and (2) after a night of sleep that ensued either 30 min or 4 h after training. The results show an early boost in performance 30 min after training and a subsequent decay across the 4 h wake interval. When sleep followed 30 min after training, post-sleep performance was stabilized at the early boost level. Sleep at 4 h after training restored performance to the early boost level, such that, 12 h after training, performance was comparable regardless of whether sleep occurred 30 min or 4 h after training. These findings indicate that sleep does not enhance but rather stabilizes motor sequence performance without producing additional gains.

Key words: consolidation; enhancement; finger sequence tapping; motor learning; sleep; stabilization

Introduction

According to a prominent model, the consolidation of procedural memories involves two successive stages: a time-dependent stabilization process that develops over wakefulness and permits the maintenance of the newly acquired skill, followed by an enhancement phase that requires sleep and induces an additional gain in performance (Walker, 2005). In this case, "enhancement" is generally understood as an increase of performance after sleep that exceeds the performance level seen at any time before sleep. Evidence for this model comes mainly from studies using the finger-tapping task, an explicit motor sequence learning task, in which participants become increasingly skilled in repeatedly tapping a specific sequence. Performance on this task was found to stabilize at the post-learning level after a period of wakefulness, whereas performance further increased after a night of sleep (Fischer et al., 2002, 2005; Walker et al., 2002, 2003b; Korman et al., 2003, 2007).

The "two-stage model" has been challenged recently by a series of divergent findings. A distinct enhancement of performance in the finger-tapping task was found to emerge already after short rest periods without intervening sleep. This "early boost" of performance has been observed consistently 5–30 min after the end of learning in the finger-tapping task (Hotermans et al., 2006, 2008; Brawn et al., 2010; Debarnot et al., 2011), as well as in other types of motor learning tasks, such as a probabilistic serial reaction time task (Schmitz et al., 2009), an implicit oculo-motor sequence learning task (Albouy et al., 2006, 2008), the pursuit rotor task (Eysenck and Frith, 1977), and even in a motor imagery task (Debarnot et al., 2011). The early boost of performance decays over the next 4–12 h during wakefulness but can be rescued by sleep (Hotermans et al., 2006; Brawn et al., 2010). Taking this early boost into account, Brawn et al. (2010) showed that sleep either stabilizes or reinstates performance but does not provide an additional gain that goes beyond the performance level observed 5 min after the end of training. Rickard et al. (2008), after controlling for confounding factors such as fatigue and reactive inhibition in the finger-tapping task, likewise observed stabilization rather than enhancement of performance across sleep. Although together these studies suggest that sleep may not enhance motor sequence performance, they remain inconclusive with regard to several important issues. Rickard et al. (2008) did not assess the early boost performance before sleep, leaving open the relation between rapid post-learning performance increases as a function of time and improvements as a result of sleep. Although Hotermans et al. (2006) and Brawn et al. (2010) did assess early boost performance, they did so by introducing an additional testing session before sleep, which basically constitutes relearning that affects consolidation (Fig. 1). Finally, none of the previous studies controlled for the timing between
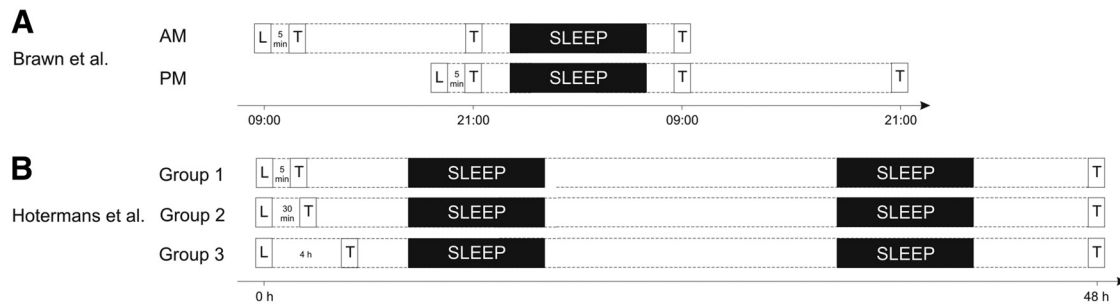
A
Brawn et al.

B
Hotermans et al.

Figure 1. Experimental design of previous studies. **A**, In the study by Brawn et al. (2010), one group (AM) learned the finger-tapping task in the morning (L) and was tested 5 min after the end of learning (T). A second and third test took place after a 12 h wake interval and after another 12 h interval including a night of sleep. The second group (PM) learned in the evening and was tested 5 min after learning, as well as after a 12 h interval including sleep and after another 12 h wake interval. The exact timing between testing before sleep and sleep onset is unclear. Performance increased significantly 5 min after learning in both groups. In the AM group performance decayed across the 12 h wake interval and was reinstated after the 12 h sleep interval. In the PM group, performance was stabilized at the 5 min level after the sleep interval and remained high after the ensuing wake interval. **B**, In the study by Hotermans et al. (2006), three groups learned the finger-tapping task (L) and were tested (T) after 5 min (Group 1), 30 min (Group 2), or 4 h (Group 3). All groups were tested a second time after 48 h. The exact time during the day when learning and testing took place as well as the timing between testing and sleep is unclear. Performance increased 5 min and 30 min but not 4 h after learning. After 48 h, performance was stabilized at the 5 min and 30 min levels in Group 1 and Group 2 and increased to the same level in Group 3. (Both Brawn et al. and Hotermans et al. report data on more experimental groups, but only the groups directly relevant to the present study are illustrated here).
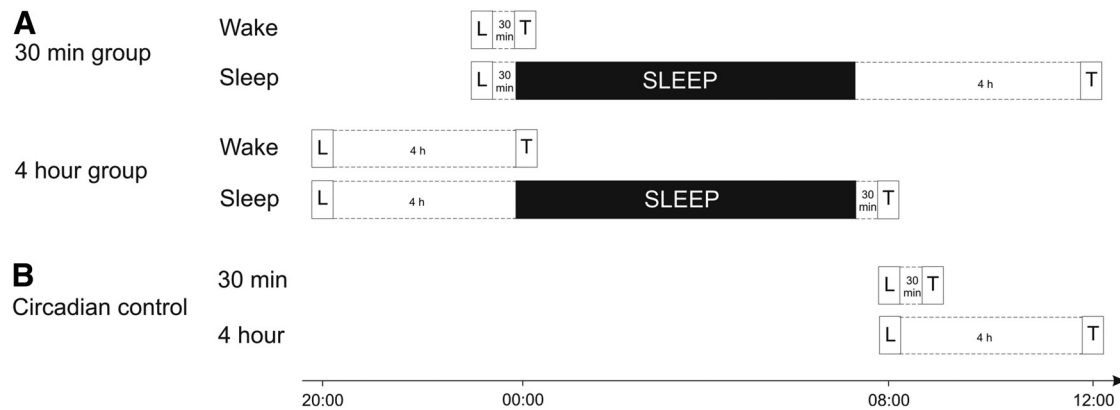
A
30 min group

4 hour group

B
Circadian control

Figure 2. Experimental design. **A**, In the main experiment, subjects participated in either the 30 min group or the 4 h group. In both groups, subjects took part in a wake condition and a sleep condition in a counterbalanced order. During learning (L), all subjects practiced on the finger sequence tapping task. In the 30 min group, subjects spent a 30 min wake interval after learning. After this interval, subjects were either tested (T) in the wake condition or they were allowed to sleep for 1 night and were tested the next day in the sleep condition. For the 4 h group, learning was followed by a 4 h wake period, after which subjects were either tested (wake condition) or went to bed for a night of sleep (sleep condition), with testing taking place the next morning. **B**, To control for circadian influences, two additional groups of subjects learned the finger sequence tapping task in the morning and were tested after 30 min or 4 h of wakefulness.

early boost assessment and sleep, leaving open whether the effect of sleep varies with the amount of time elapsed between learning, early boost, and sleep.

Here we examined the role of sleep for the consolidation of sequential finger tapping by (1) taking the early performance boost into account, (2) controlling for relearning before sleep when assessing boost performance, and (3) directly manipulating the time window between learning, early boost, and sleep. We hypothesized that performance shows an early boost 30 min but not 4 h after training. Furthermore, we expected sleep ensuing 30 min after training to stabilize performance, whereas sleep ensuing 4 h after training was assumed to reinstate performance of the early boost level.

Materials and Methods

Participants. Thirty-one healthy nonsmoking subjects (15 females) between the ages of 19 and 32 years (mean ± SD age, 23.94 ± 3.02 years) participated in the main experiment. Half of the participants were randomly assigned to a group with a 30 min wake interval between learning and sleep (30 min group; $n = 16$), and the other half stayed awake for 4 h after learning before they were allowed to sleep (4 h group; $n = 15$; Fig. 2A). As a control for circadian influences, another 30 healthy nonsmok-

ing subjects (mean ± SD age, 22.87 ± 3.23 years; range, 18–30 years; eight females) stayed awake for 30 min ($n = 15$) or 4 h ($n = 15$) after learning in the morning (Fig. 2B). All participants reported to have no history of neurological, psychiatric, or sleep disorders and did not take any medication at the time of the study. None of the subjects did shift work or night work for at least 3 weeks before the experiments. Participants were instructed to ingest no alcohol and caffeine and not to take any naps during the experimental days. In preparation for experimental nights, all subjects spent one adaptation night in the sleep laboratory under experimental conditions. The study was approved by the local ethics committee of the University of Lübeck. All participants provided written informed consent and were paid for participation.

Design and experimental protocol. In the main experiment, subjects took part in either the 30 min group or the 4 h group. The duration of these intervals was chosen based on evidence from Hotermans et al. (2006) showing that performance in the finger-tapping task increases significantly 30 min after training and returns back to training levels 4 h after training. In both groups, subjects participated in a short wake condition and a longer sleep condition according to a counterbalanced within-subject design (Fig. 2A). For each subject, both conditions were separated by at least 2 weeks.

In the 30 min wake condition, the learning session started at 11:00 P.M. Subjects first performed a vigilance task and then trained on the

finger sequence tapping task. After learning, subjects spent a short wake interval of 30 min during which they watched parts of a movie. The testing session took place thereafter at ~12:00 A.M. In the 30 min sleep condition, participants arrived at the laboratory at 10:00 P.M., and electrodes were attached for sleep recordings. The learning session for the 30 min sleep condition was identical to the 30 min wake condition, with the only exception that, after the 30 min wake interval, subjects were not tested on the finger-tapping task but went to bed at 12:00 A.M., immediately after watching parts of the movie. Subjects were allowed to sleep normally and were awakened at 7:30 A.M. the next morning. After awakening, participants continued watching movies for 4 h and were tested on the finger sequence tapping task at ~12:00 P.M.. After performance on the finger sequence tapping task, participants completed the vigilance task again.

In the 4 h wake condition, the learning session started at 7:30 P.M. but was otherwise identical to the 30 min wake condition. After learning, participants watched a defined set of movies for 4 h, and the testing session took place at 12:00 A.M. In the 4 h sleep condition, participants arrived at the laboratory at 7:00 P.M. and were attached to the electrodes for sleep recordings. Similar to the wake condition, the learning session started at 7:30 P.M., and participants watched a set of movies for 4 h after learning. Subjects were not tested on the finger-tapping task in the evening but went to bed at 12:00 A.M. They were awakened at 7:30 A.M. the next morning, and the testing session took place ~30 min after awakening at 8:00 A.M.

In the circadian control groups, subjects trained on the finger sequence tapping task at 8:00 A.M. After learning, subjects watched movies for either 30 min in the 30 min group or 4 h in the 4 h group and were tested thereafter (Fig. 2B).

Finger-tapping task. The finger sequence tapping task represents a widely used procedural memory task to assess explicit motor skill learning (Walker et al., 2002, a,b; Rasch et al., 2009). Subjects were instructed to press four numeric keys on a standard computer keyboard with the fingers (except thumb) of their nondominant hand, repeating a five-element sequence as fast and as accurately as possible without interruption. The numeric sequence was displayed at the top of the screen at all times to exclude any working memory component to the task. Two different sequences were used in a counterbalanced order for the sleep condition and the wake condition in the 30 min and 4 h groups (sequence A, 4-2-3-1-4; sequence B, 4-1-3-2-4). Counterbalancing sequences and conditions was particularly important because practice on the first-trained sequence tended to transfer to the second-trained sequence, with performance on the second-trained sequence being on average one correct sequence faster ($p < 0.10$, averaged across both groups). Counterbalancing sequences and conditions ensured that, on average, any practice effects evened out.

Before performing the experimental sequence, subjects tapped a warm-up sequence four times to get used to the task (sequence, 1-1-2-3-4). Immediately after the warm-up session, subjects were asked whether they understood the task, and, after a positive answer, the actual training session started. The training session consisted of 12 contiguous blocks of 30 s each, followed by 30 s of rest. The test session consisted of three contiguous blocks of 30 s each, followed by 30 s of rest. Each key press produced a star on the screen, forming a row from left to right, to indicate the present location in the sequence without providing accuracy feedback. Performance was determined as the number of correctly completed sequences per block, reflecting a combined measure of speed and accuracy. Additionally, mean reaction times (in milliseconds) and error rate (number of errors relative to total number of tapped sequences) were obtained.

The average performance on the last three blocks of the learning session served as a measure of learning performance, and the average performance of the three test trials was used as a measure of test performance. Performance change was assessed as the percentage difference between test performance and learning performance, with learning performance set to 100%. Because we omitted the assessment of pre-sleep performance level in the sleep conditions (to avoid relearning), we estimated pre-sleep performance in the sleep conditions from performance change values of the respective wake conditions. To this end, we took the percentage performance change of the wake condition for each

individual subject and added this value to the learning performance of the sleep condition. For example, if a subject improved performance from learning to test by 20% in the 30 min wake condition, we added 20% to the learning performance in the 30 min sleep condition to obtain the estimated pre-sleep performance level. The performance change from pre-sleep to post-sleep testing was then determined as the percentage difference between test performance and estimated pre-sleep performance level.

Vigilance task. To assess general alertness, subjects performed a vigilance task twice: once before learning and a second time after testing. Participants were required to respond as fast as possible to the appearance of a red dot on a computer screen. The dot appeared every 2–10 s on the left or right side of the screen, and subjects had to press the corresponding left or right button. The task lasted ~10 min.

Sleep data. Polysomnography included electroencephalographic (EEG), electromyographic (EMG), and electrooculographic (EOG) recordings. For EEG recordings, nine electrodes were placed on the scalp (at positions F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4 according to the international 10–20 system), with two reference electrodes over the mastoids. EMG recordings were obtained from two electrodes placed on the chin, and for EOG recordings, two electrodes were placed on both outer canthi. Polysomnographic recordings were visually scored offline by two experienced scorers according to standard criteria as wake, sleep stages 1, 2, 3, and 4 (with sleep stages 3 and 4 combined for slow-wave sleep), and rapid eye movement (REM) sleep (Rechtschaffen and Kales, 1968).

For a more fine-grained analysis of changes in sleep parameters after learning, power spectral analysis was applied. All epochs with artifacts and arousals were rejected automatically (movement artifacts in the EMG exceeding ± 50 μV). Artifact- and arousal-free epochs of non-REM sleep stages 2, 3, and 4 (according to the sleep scores) were then separated into blocks of 2048 data points each (~10.2 s), with an overlap of 205 data points between blocks. Power density was calculated using fast Fourier transformation after applying a tapered Hanning window to each block. Mean power density in the following frequency bands was determined: slow oscillations (0.7–1.2 Hz), delta (1.2–4 Hz), slow spindles (9–12 Hz), and fast spindles (12–15 Hz). (The results remain the same with the frequency range of 0.5–1 Hz for slow oscillations and 1–4 Hz for delta.) Data were analyzed separately for electrode positions F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4 and averaged across non-REM sleep of the entire sleep period. Additionally, the first 20 min of non-REM sleep were analyzed separately because it is known that changes in sleep parameters after learning are most pronounced during the first few minutes of non-REM sleep (Huber et al., 2004). The period length of 20 min was chosen based on a previous study, which successfully applied the 20 min time window (Wilhelm et al., 2011).

Statistical analyses. In the main experiment, data of the finger sequence tapping task were first analyzed using repeated-measures ANOVA with the within-subject factor sleep/wake and the between-subject factor 30 min/4 h. For analysis of performance changes across the retention interval and for analysis of the vigilance task, the additional within-subject factor learn/test was introduced to the ANOVA. For comparison of data from the main experiment with the circadian control groups, the additional between-subject factor evening/morning was used. In a second step, to account for the estimated pre-sleep performance level, another repeated-measures ANOVA was applied with the within-subject factor learn/pre-sleep/test and the between-subject factor 30 min sleep/4 h sleep. In case of significant ANOVA effects, planned *post hoc* one-way ANOVAs and unpaired and paired samples *t* tests were applied. Sleep data were analyzed using unpaired *t* tests. Correlations were calculated according to Pearson's correlation coefficients. Level of significance was set to $p = 0.05$. Greenhouse–Geisser correction for degrees of freedom was applied when appropriate.

In the 4 h group, data of one subject had to be discarded because of missing data. In the 4 h circadian control group, three subjects had to be excluded from analysis of the vigilance task because of technical problems. One subject from the 4 h group had to be excluded from power spectral analyses because of bad quality of the EEG signal. Data from another four subjects were excluded for single electrode sites if electrodes were lost. Overall, the number of subjects available for analysis for single
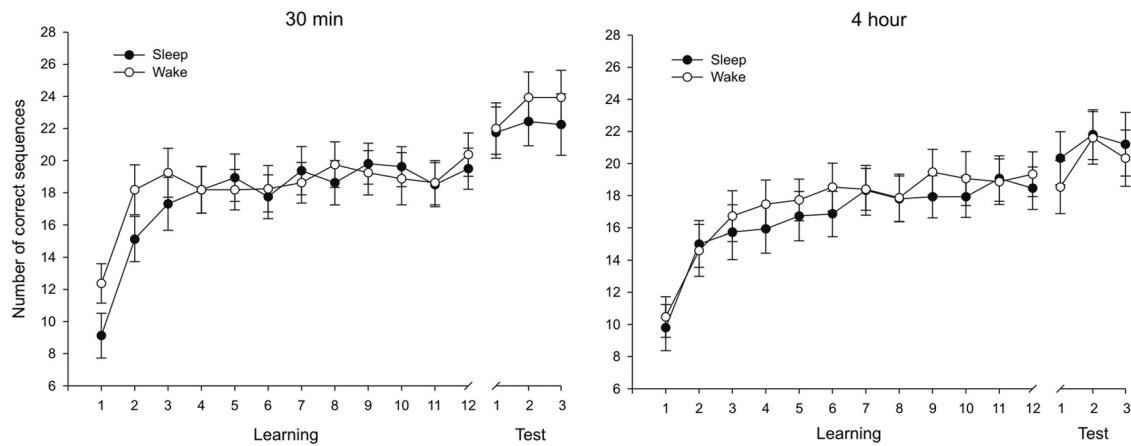
Figure 3. Learning curves of the finger sequence tapping task. The number of correct sequences tapped during each of the 12 learning blocks and three test blocks is illustrated for the wake and sleep conditions of the 30 min group (left) and the 4 h group (right), respectively. Note that subjects reached asymptotic performance during the last six learning blocks in all groups and conditions. Mean performance during the final three learning blocks (10, 11, and 12) did not differ significantly from mean performance during the three preceding blocks (7, 8, and 9; all p values >0.35). Means ± SEMs are shown.
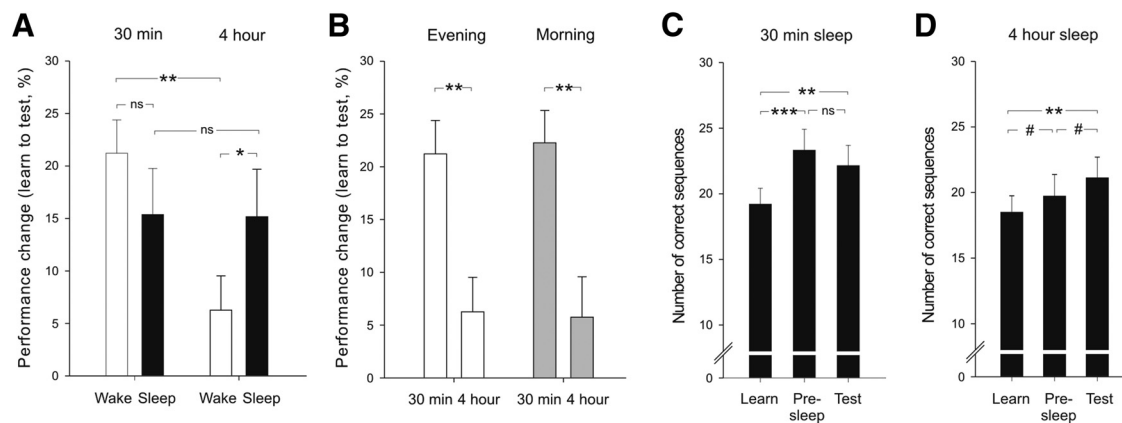

Figure 4. Finger sequence tapping performance. **A**, The early boost in performance that was expressed after 30 min of wakefulness was preserved at delayed testing if sleep followed 30 min after learning. Compared with the improvement observed 30 min after learning, performance decreased significantly over the 4 h wake interval. If subjects were allowed to sleep after 4 h of wakefulness, performance was restored to the level seen 30 min after learning. **B**, The circadian control groups (Morning, gray bars) showed the same pattern of improvement as the respective wake groups from the main experiment (Evening, white bars), with performance decreasing from 30 min to 4 h after learning. **C, D**, Absolute number of correctly tapped sequences during learning (Learn) and testing (Test) and the estimated performance level before sleep (Pre-sleep) in the 30 min and 4 h sleep conditions. Means ± SEMs are shown. ns, Not significant. $^{\#}p < 0.06$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

electrode sites ranged from 13 to 14 in the 4 h group and 13 to 16 in the 30 min group.

Results

Finger sequence tapping task

We first analyzed performance changes from learning to testing in the number of correctly tapped sequences, without considering the estimated pre-sleep level. Learning curves are illustrated in Figure 3. As expected, finger-tapping performance increased distinctly 30 min after training. When subjects were tested 30 min after the end of learning in the 30 min wake condition, they showed a highly significant improvement of 21.22 ± 3.17% compared with the end of learning ($t_{(15)} = 6.34$, $p < 0.001$; Fig. 4A). This improvement was preserved when subjects were allowed to sleep after the 30 min wake interval in the 30 min sleep condition. When tested after sleep, subjects improved their performance by 15.39 ± 4.37% compared with learning performance ($t_{(15)} = 3.57$, $p = 0.003$). The improvement observed after sleep did not differ significantly from the early boost performance 30 min after training (30 min sleep vs 30 min wake: $t_{(15)} = 1.08$, $p = 0.30$). After 4 h of wakefulness in the 4 h wake condition, subjects showed only a

marginal performance increase of 6.26 ± 3.27% from learning to test ($t_{(14)} = 2.05$, $p = 0.06$). When subjects were allowed to sleep after the 4 h wake period in the 4 h sleep condition, performance improved significantly by 15.16 ± 4.52% ($t_{(14)} = 3.31$, $p = 0.005$), with this improvement being larger than that observed after the 4 h wake interval (4 h sleep vs 4 h wake: $t_{(14)} = 2.28$, $p = 0.039$).

Comparing the improvement after 30 min of wakefulness with that after the 4 h wake interval revealed a significant decrease in performance from 30 min to 4 h. The early performance boost of 21.22 ± 3.17% after 30 min dropped to 6.26 ± 3.27% over the course of 4 h (30 min wake vs 4 h wake: $t_{(29)} = 3.29$, $p = 0.003$). Sleep after the 4 h wake interval resulted in an improvement similar to the improvement seen in the 30 min sleep condition at the end of the 12 h retention interval (4 h sleep: 15.16 ± 4.52%; 30 min sleep: 15.39 ± 4.37%; $t_{(29)} = 0.04$, $p = 0.97$; interaction of 30 min/4 h × sleep/wake: $F_{(1,29)} = 4.76$, $p = 0.037$).

To exclude that the sharp decline of performance improvement from 30 min to 4 h after learning resulted from circadian confounds, for example from subjects being more prone to forgetting and/or unspecific interference during the evening hours,

Table 1. Performance on the finger sequence tapping task

	Number of correct sequences		Reaction time		Error rate	
	Learning	Test	Learning	Test	Learning	Test
30 min group						
Wake	19.29 ± 1.15	**23.29 ± 1.45**	258.43 ± 21.69	**217.31 ± 22.30**	0.19 ± 0.05	0.14 ± 0.04
Sleep	19.21 ± 1.05	**22.15 ± 1.42**	268.59 ± 21.35	**228.13 ± 19.05**	0.12 ± 0.03	0.13 ± 0.03
4 h group						
Wake	19.09 ± 1.60	*20.16 ± 1.69*	271.71 ± 22.40	267.48 ± 23.03	0.19 ± 0.04	0.14 ± 0.03
Sleep	18.49 ± 1.42	**21.11 ± 1.71**	281.12 ± 22.05	**247.76 ± 19.67**	0.19 ± 0.04	0.17 ± 0.03
Circadian control						
30 min	18.67 ± 0.97	**22.82 ± 1.27**	265.01 ± 21.97	**219.52 ± 22.11**	0.20 ± 0.05	0.18 ± 0.06
4 h	18.44 ± 1.21	19.36 ± 1.26	283.36 ± 21.97	274.19 ± 22.11	0.11 ± 0.03	0.10 ± 0.02

Means ± SEMs are shown. Bold and italic numbers denote significant changes from learning to test (results of pairwise comparisons; bold, $p < 0.01$; italic, $p = 0.049$). Reaction time is shown in milliseconds and error rate as number of errors per tapped sequence.

we compared performance of the 30 min and 4 h wake groups of the main experiment with the 30 min and 4 h circadian control groups. Like in the main experiment, there was a significant decline in performance from 30 min to 4 h in the circadian control groups ($t_{(28)} = 3.37, p = 0.002$) that was independent of circadian phase ($F_{(1,57)} = 22.18, p = 0.001$ for main effect of 30 min/4 h; $p > 0.80$ for main effect of evening/morning and interaction effect of 30 min/4 h × evening/morning; Fig. 4B).

Finally, we analyzed the time course of performance changes in the sleep conditions separately while taking the estimated pre-sleep performance level into account. The estimated pre-sleep performance level was calculated by adding each individual subject's percentage performance change of the wake condition to the learning performance of the respective sleep condition (for details, see Materials and Methods). In this way, we were able to analyze the time course of performance changes from learning to pre-sleep level to testing after sleep. This analysis revealed distinctly different patterns depending on whether sleep ensued 30 min or 4 h after the end of training ($F_{(2,58)} = 3.69, p = 0.031$ for interaction of learn/pre-sleep/test × 30 min sleep/4 h sleep). In the 30 min sleep condition, performance increased distinctly from learning to pre-sleep ($t_{(15)} = 5.42, p < 0.001$) and remained stable from pre-sleep to post-sleep test performance ($t_{(15)} = 1.00$, $p = 0.33; F_{(2,30)} = 9.80, p < 0.001$ for main effect of learn/pre-sleep/test; Fig. 4C), whereas in the 4 h sleep condition, performance only tended to increase from learning to pre-sleep ($t_{(14)} = 2.10, p = 0.055$) and tended to increase again from pre-sleep to testing after sleep ($t_{(15)} = 2.12, p = 0.053; F_{(2,28)} = 8.62, p = 0.001$ for main effect of learn/pre-sleep/test; Fig. 4D). To control for possible effects of the order of conditions, i.e., whether subjects first participated in the sleep or the wake condition, we included the additional factor order in the ANOVA. As in the main analysis, the interaction effect of learn/pre-sleep/test × 30 min sleep/4 h sleep remained significant ($F_{(2,54)} = 3.68, p = 0.039$). Importantly, none of the order effects reached significance (all p values > 0.20), confirming that the results were not affected by the order of the sleep and wake conditions.

Initial learning of the finger-tapping task was comparable between the 30 min group and the 4 h group, as well as between the wake and sleep conditions in the main experiment (all p values > 0.50; Table 1). The two circadian control groups likewise showed comparable learning performance ($p = 0.89$), and learning performance in these groups did not differ from learning performance in the 30 min and 4 h wake groups of the main experiment (all p values > 0.60 for main effects and interaction of 30 min/4 h × evening/morning). The reaction time data closely mirrored the effects obtained with the number of correctly tapped sequences, whereas error rate did not show any significant differences (Table 1).

Table 2. Sleep stage distribution

	30 min group	4 h group	P
TST (min)	428.50 ± 6.61	432.70 ± 3.49	>0.50
Wake (%)	2.98 ± 1.44	0.76 ± 0.38	>0.15
S1 (%)	0.97 ± 0.32	1.04 ± 0.20	>0.80
S2 (%)	58.21 ± 2.12	61.45 ± 1.82	>0.25
SWS (%)	15.29 ± 0.87	13.53 ± 1.75	>0.35
REM sleep (%)	20.88 ± 0.71	20.98 ± 1.12	>0.90

Means ± SEMs are shown. Total sleep time (TST) is indicated in minutes with sleep stage 1 (S1), sleep stage 2 (S2), slow-wave sleep (SWS, combined sleep stages 3 and 4), and REM sleep as percentage of total sleep time for the sleep condition of the 30 min and 4 h groups, respectively. p values refer to pairwise group comparisons.

Table 3. Performance on the vigilance task

	Reaction time	
	Learning	Test
30 min group		
Sleep	371.23 ± 10.69	374.94 ± 11.79
Wake	366.25 ± 09.58	397.64 ± 11.61
4 h group		
Sleep	373.24 ± 11.04	376.67 ± 12.17
Wake	358.92 ± 09.89	376.97 ± 11.99
Circadian control		
30 min	351.15 ± 12.31	369.03 ± 15.78
4 h	354.54 ± 08.12	364.73 ± 08.22

Means ± SEMs are shown. Reaction time is shown in milliseconds.

Sleep data

In both the 30 min and the 4 h sleep conditions, the subjects showed normal sleep patterns. The groups did not differ in total sleep time or time spent in single sleep stages (all p values > 0.15; Table 2). Moreover, none of the power density measures at any of the electrode sites differed between groups, for either the whole non-REM sleep period or the first 20 min of non-REM sleep (all p values > 0.05). To explore whether finger sequence tapping performance, i.e., performance change from pre-sleep to post-sleep test, was related to any of the sleep parameters, we ran correlational analyses separately for the 30 min sleep condition and the 4 h sleep condition. None of the correlations reached significance after correcting for multiple comparisons.

Vigilance task

The 30 min and 4 h groups and the sleep and wake conditions were comparable in general alertness as measured by mean reaction time in the vigilance task ($p > 0.10$ for main effects of 30 min/4 h and sleep/wake, as well as for the interaction of 30 min/4 h × sleep/wake; Table 3). Also in the circadian control groups, there were no differences in reaction time between the 30 min and the 4 h group ($p > 0.40$ for main effect of 30 min/4 h). Reaction

times in the vigilance task were not associated with performance in the finger-tapping task in any of the groups or conditions at either learning or testing (all *p* values >0.10).

Discussion

The present results show that sleep stabilizes finger sequence tapping skill but does not enhance performance. We found a distinct early performance boost in finger tapping 30 min after training that decayed over the course of 4 h of wakefulness. Sleep either (1) stabilized performance at the early boost level if it followed 30 min after training or (2) restored the early boost performance level if sleep was delayed by 4 h. These findings indicate that, contrary to assumptions of the two-stage model of procedural memory consolidation (Walker, 2005), sleep does not enhance procedural task performance in the sense of "true gains," i.e., gains beyond the performance level expressed at any time before sleep.

The notion that motor sequence performance is enhanced by sleep is widely held in the field of procedural memory research, with "enhancement" meaning that sleep produces additional gains in performance in the absence of any additional practice. The most prominent model incorporating this view is the two-stage model of procedural memory consolidation, assuming that motor sequence performance becomes stabilized during wakefulness and enhanced during sleep (Walker, 2005). It was only recently that first studies provided contrary evidence putting this assumption into question (Hotermans et al., 2006; Rickard et al., 2008; Brawn et al., 2010). Hotermans et al. (2006) showed that performance in finger tapping already strongly increases 5–30 min after training, suggesting that enhancements in motor sequence performance do not necessarily require sleep. Brawn et al. (2010) replicated the early boost in tapping performance 5 min after training and found that sleep either maintains or restores performance of the early boost level.

Although compellingly designed and suggestive of a stabilizing effect of sleep, these previous studies were not entirely conclusive as to the central question whether actual sleep-dependent enhancements in motor sequence performance are possible. In the studies by Hotermans et al. (2006) and Brawn et al., 2010, participants were tested twice on the finger-tapping task: once before the sleep/wake period to determine the early boost level and a second time after the intervals of sleep or wakefulness. The additional testing session before the retention interval might have immediately strengthened the memory representation and/or affected subsequent sleep-dependent consolidation processes. In the declarative memory domain, it is well known that retrieval sessions represent powerful means to strengthen the memory trace, even more so than actual relearning sessions (Roediger and Karpicke, 2006; Karpicke and Roediger, 2008). For procedural memory, this is even more obvious because any testing of performance (which necessarily entails performing the task) represents additional practice of the skill. Therefore, it is unclear whether post-sleep performance in the studies by Brawn et al., 2010 and Hotermans et al. (2006) can be attributed to a stabilizing/restoring effect of sleep or rather to the additional practice during testing.

In the present study, we avoided additional practice by omitting the testing session before sleep while still assessing pre-sleep performance in the same subjects in a separate condition. We show that, even when potential relearning effects are excluded, sleep preserves or reinstates early boost performance but does not enhance performance. Sleep stabilizes the early boost performance if it follows 30 min after learning. In this case, sleep onset coincides with the highest pre-sleep performance level. If sleep is delayed by 4 h, the highest performance level decays over the

course of 4 h of wakefulness, such that performance is lower immediately before sleep onset. In this case, sleep increases performance and thereby reinstates the highest performance level that was expressed 30 min after training. Although in the case of reinstating performance in the 4 h group sleep increased performance compared with the immediate pre-sleep level, the final post-sleep performance level did not exceed the level of performance that was already seen during the early boost, i.e., 30 min after learning. Taking the 30 min early boost level as a reference for the actual learning performance, our findings show that sleep does not produce additional performance benefits. It is a matter of interpretation whether the reinstatement of an earlier performance level in the 4 h group can be called "stabilization" or whether it can be regarded as some form of enhancement. Importantly, however, performance after sleep never exceeded performance levels that were already attained at some point before sleep, strongly arguing against the notion that sleep produces "true gains" in the absence of additional practice.

The early boost represents a substantially higher level of performance than that observed at the end of training. Although the mechanisms underlying this phenomenon are unknown, some diverging explanations have been put forward. Early studies investigating reminiscence suggested that the early boost represents a form of consolidation (Eysenck and Frith, 1977). Others have suggested that the early boost originates from inhibitory fatigue that builds up during performance and is released after a short period of rest (Heuer and Klein, 2003). Alternatively, the early boost might be a consolidation precursor that primes subsequent long-term consolidation processes, or it might represent the active/labile state of a motor memory as conceptualized in the reconsolidation framework (Hotermans et al., 2006; Schmitz et al., 2009). Studies applying transcranial magnetic stimulation implicated a role of the primary motor cortex (M1) in the emergence of the early boost. Interfering with M1 during the rest period after learning diminished or even abolished the early boost (Hotermans et al., 2008; Debarnot et al., 2011). Other findings suggest that the early boost is characterized by a transient facilitation of cortical processing as evidenced by an enhancement of the N100 amplitude and a reduction in the P300 latency of event-related potentials (Schmitz et al., 2009). The relevance of these physiological mechanisms and their relation to sleep-associated consolidation processes will have to be tested systematically in future studies. Importantly, the interpretation of the present findings is entirely independent of the mechanisms underlying the early boost effect.

It has been a matter of debate whether the amount of time elapsing between learning and sleep onset affects the consolidation benefits of sleep for procedural memories. We show here that, for motor sequence consolidation in the finger-tapping task, sleep delayed by 4 h is as effective as sleep 30 min after learning. Although performance decayed from 30 min to 4 h after training, sleep restored performance in the 4 h condition such that, 12 h after training, the performance level was identical for sleep that ensued 30 min and 4 h after training. Additional circadian control groups ensured that the decay observed from 30 min to 4 h was not attributable to differences in circadian time, as evidenced by a similar time course of the early boost and the subsequent decay after training in the morning and the evening. Sleepiness and vigilance were ruled out as confounding factors by showing that reaction times in the vigilance task were comparable between groups and were not associated with finger-tapping performance. Moreover, finger-tapping performance was comparable between the 4 h and the 30 min sleep conditions at testing after the 12 h interval, suggesting that sleepiness and alertness did not affect performance.

Considering the present findings, future studies should assess the early boost after training when studying performance improvements across sleep. Our findings question the validity of measures determining sleep-dependent improvements as increases from the end of learning to testing after sleep. We show here that performance at the end of learning does not reflect the actual pre-sleep skill level. Consequently, the improvement measure that has been used widely in sleep studies does not represent a measure of sleep-dependent motor consolidation processes. There is, in fact, a principal problem in obtaining a "real" performance measure before sleep because each testing session constitutes a relearning session that potentially alters the consolidation process. Here we provide an alternative method to assess pre-sleep performance by omitting the testing session before sleep and, instead, estimating pre-sleep performance based on a different experimental condition in the same subjects. The finding that two additional independent control groups closely replicated the early performance boost 30 min and 4 h after training suggests that this estimate is sufficiently reliable.

In the future, the present findings should be extended to other tasks and paradigms in the procedural memory domain. Although the finger sequence tapping task has been most frequently used to study the role of sleep in the offline consolidation of motor sequence skills (Walker et al., 2003a,b; Wilhelm et al., 2008; Rasch et al., 2009; Dresler et al., 2010; Genzel et al., 2012), this task does not allow for dissociating mere improvements in general button-pressing ability from specific sequence learning. Thus, we cannot exclude that, in the present study, sleep provided additional benefits (e.g., more anticipation of the sequence) that are not measurable with the current paradigm. To test for this possibility, paradigms, such as the serial reaction time task, that allow for a comparison between sequential and random button presses should be applied (Robertson et al., 2004; Song and Cohen, 2014).

References

Albouy G, Ruby P, Phillips C, Luxen A, Peigneux P, Maquet P (2006) Implicit oculomotor sequence learning in humans: time course of offline processing. Brain Res 1090:163–171. CrossRef Medline

Albouy G, Sterpenich V, Balteau E, Vandewalle G, Desseilles M, Dang-Vu T, Darsaud A, Ruby P, Luppi PH, Degueldre C, Peigneux P, Luxen A, Maquet P (2008) Both the hippocampus and striatum are involved in consolidation of motor sequence memory. Neuron 58:261–272. CrossRef Medline

Brawn TP, Fenn KM, Nusbaum HC, Margoliash D (2010) Consolidating the effects of waking and sleep on motor-sequence learning. J Neurosci 30:13977–13982. CrossRef Medline

Debarnot U, Clerget E, Olivier E (2011) Role of the primary motor cortex in the early boost in performance following mental imagery training. PLoS One 6:e26717. CrossRef Medline

Dresler M, Kluge M, Genzel L, Schüssler P, Steiger A (2010) Impaired offline memory consolidation in depression. Eur Neuropsychopharmacol 20:553–561. CrossRef Medline

Eysenck HJ, Frith CD (1977) Reminiscence, motivation, and personality: a case study in experimental psychology. New York: Plenum.

Fischer S, Hallschmid M, Elsner AL, Born J (2002) Sleep forms memory for finger skills. Proc Natl Acad Sci U S A 99:11987–11991. CrossRef Medline

Fischer S, Nitschke MF, Melchert UH, Erdmann C, Born J (2005) Motor memory consolidation in sleep shapes more effective neuronal representations. J Neurosci 25:11248–11255. CrossRef Medline

Genzel L, Kiefer T, Renner L, Wehrle R, Kluge M, Grözinger M, Steiger A, Dresler M (2012) Sex and modulatory menstrual cycle effects on sleep related memory consolidation. Psychoneuroendocrinology 37:987–998. CrossRef Medline

Heuer H, Klein W (2003) One night of total sleep deprivation impairs implicit learning in the serial reaction task, but not the behavioral expression of knowledge. Neuropsychology 17:507–516. CrossRef Medline

Hotermans C, Peigneux P, Maertens de Noordhout A, Moonen G, Maquet P (2006) Early boost and slow consolidation in motor skill learning. Learn Mem 13:580–583. CrossRef Medline

Hotermans C, Peigneux P, de Noordhout AM, Moonen G, Maquet P (2008) Repetitive transcranial magnetic stimulation over the primary motor cortex disrupts early boost but not delayed gains in performance in motor sequence learning. Eur J Neurosci 28:1216–1221. CrossRef Medline

Huber R, Ghilardi MF, Massimini M, Tononi G (2004) Local sleep and learning. Nature 430:78–81. CrossRef Medline

Karpicke JD, Roediger HL 3rd (2008) The critical importance of retrieval for learning. Science 319:966–968. CrossRef Medline

Korman M, Raz N, Flash T, Karni A (2003) Multiple shifts in the representation of a motor sequence during the acquisition of skilled performance. Proc Natl Acad Sci U S A 100:12492–12497. CrossRef Medline

Korman M, Doyon J, Doljansky J, Carrier J, Dagan Y, Karni A (2007) Daytime sleep condenses the time course of motor memory consolidation. Nat Neurosci 10:1206–1213. CrossRef Medline

Rasch B, Pommer J, Diekelmann S, Born J (2009) Pharmacological REM sleep suppression paradoxically improves rather than impairs skill memory. Nat Neurosci 12:396–397. CrossRef Medline

Rechtschaffen A, Kales A (1968) A manual of standardized terminology, techniques, and scoring system for sleep stages of human subjects. Bethesda, MD: Department of Health, Education, and Welfare/National Institutes of Health.

Rickard TC, Cai DJ, Rieth CA, Jones J, Ard MC (2008) Sleep does not enhance motor sequence learning. J Exp Psychol Learn Mem Cogn 34:834–842. CrossRef Medline

Robertson EM, Pascual-Leone A, Miall RC (2004) Current concepts in procedural consolidation. Nat Rev Neurosci 5:576–582. CrossRef Medline

Roediger HL, Karpicke JD (2006) Test-enhanced learning: taking memory tests improves long-term retention. Psychol Sci 17:249–255. CrossRef Medline

Schmitz R, Schabus M, Perrin F, Luxen A, Maquet P, Peigneux P (2009) Recurrent boosting effects of short inactivity delays on performance: an ERPs study. BMC Res Notes 2:170. CrossRef Medline

Song S, Cohen LG (2014) Practice and sleep form different aspects of skill. Nat Commun 5:3407. CrossRef Medline

Walker MP (2005) A refined model of sleep and the time course of memory formation. Behav Brain Sci 28:51–64; discussion 64–104.

Walker MP, Brakefield T, Morgan A, Hobson JA, Stickgold R (2002) Practice with sleep makes perfect: sleep-dependent motor skill learning. Neuron 35:205–211. CrossRef Medline

Walker MP, Brakefield T, Hobson JA, Stickgold R (2003a) Dissociable stages of human memory consolidation and reconsolidation. Nature 425:616–620. CrossRef Medline

Walker MP, Brakefield T, Seidman J, Morgan A, Hobson JA, Stickgold R (2003b) Sleep and the time course of motor skill learning. Learn Mem 10:275–284. CrossRef Medline

Wilhelm I, Diekelmann S, Born J (2008) Sleep in children improves memory performance on declarative but not procedural tasks. Learn Mem 15:373–377. CrossRef Medline

Wilhelm I, Diekelmann S, Molzow I, Ayoub A, Mölle M, Born J (2011) Sleep selectively enhances memory expected to be of future relevance. J Neurosci 31:1563–1569. CrossRef Medline