

# Sleep Accelerates the Improvement in Working Memory Performance

Kenichi Kuriyama,<sup>1,2</sup> Kazuo Mishima,<sup>1</sup> Hiroyuki Suzuki,<sup>1</sup> Sayaka Aritake,<sup>1</sup> and Makoto Uchiyama<sup>3</sup>

Departments of <sup>1</sup>Psychophysiology and <sup>2</sup>Adult Mental Health, National Institute of Mental Health, National Center of Neurology and Psychiatry, Tokyo 187-8553, Japan, and <sup>3</sup>Department of Neuropsychiatry, Nihon University School of Medicine, Tokyo 173-8610, Japan

Working memory (WM) performance, which is an important factor for determining problem-solving and reasoning ability, has been firmly believed to be constant. However, recent findings have demonstrated that WM performance has the potential to be improved by repetitive training. Although various skills are reported to be improved by sleep, the beneficial effect of sleep on WM performance has not been clarified. Here, we show that improvement in WM performance is facilitated by posttraining naturalistic sleep. A spatial variant of the *n*-back WM task was performed by 29 healthy young adults who were assigned randomly to three different experimental groups that had different time schedules of repetitive *n*-back WM task sessions, with or without intervening sleep. Intergroup and intersession comparisons of WM performance (accuracy and response time) profiles showed that *n*-back accuracy after posttraining sleep was significantly improved compared with that after the same period of wakefulness, independent of sleep timing, subject's vigilance level, or circadian influences. On the other hand, response time was not influenced by sleep or repetitive training schedules. The present study indicates that improvement in *n*-back accuracy, which could reflect WM capacity, essentially benefits from posttraining sleep.

**Key words:** sleep; working memory capacity; memory consolidation; *n*-back task; skill learning; intelligence

## Introduction

Working memory (WM) is understood to be a cognitive system for both the temporary storage and manipulation of remembered information. It is regarded as a specific process by which a remembered stimulus is held “on-line” to guide behavior in the absence of external cues or prompts (Baddeley and Hitch, 1974; Goldman-Rakic, 1996; Owen et al., 1996). The maximum amount of information that can be retained in the WM, referred to as WM capacity, is an important factor for determining problem-solving and reasoning ability (Kyllonen and Christal, 1990; Fry and Hale, 1996; Hale et al., 1997). The WM encompasses the concept of traditional “short-term memory,” and consequently both WM and short-term memory share cognitive architecture and functional neuroanatomy. Miller (1956) reported that the capacity for WM (which is still sometimes called “short-term” memory) in healthy adults is restricted to within  $\sim 7 \pm 2$  chunks. Since then, it has been firmly believed that there exists a limit to the capacity of WM, and subsequent studies confirmed that this limit is approximately 4 items without the use of any hidden strategies (Luck and Vogel, 1997; Cowan, 2001).

The “*n*-back” procedure (Gevins and Cuttillo, 1993; Callicott et al., 1998, 1999; McEvoy et al., 1998) has been used in many

human studies to investigate the characteristics of WM performance or the neural basis of WM processes (Callicott et al., 1998; Jansma et al., 2004; Mattay et al., 2006). A very recent study has shown that the limit to WM capacity is determined by the ability to remember only relevant information, and that the prefrontal cortex and basal ganglia activities preceding the filtering of irrelevant information are associated with interindividual differences in WM capacity (McNab and Klingberg, 2008). Some studies have shown that the training of WM may lead to effects that go beyond a specific training effect (Olesen et al., 2004; Westerberg and Klingberg, 2007; Jaeggi et al., 2008). Olesen et al. (2004) presented progressive evidence obtained by functional magnetic resonance imaging that repetitive training improves spatial WM performance [both accuracy and response time (RT)] associated with increased cortical activity in the middle frontal gyrus and the superior and inferior parietal cortices. Such a finding suggests that training-induced improvement in WM performance could be based on neural plasticity, similar to that for other skill-learning characteristics.

A growing body of literature in recent years holds that sleep plays a crucial role in the development of skill learning. Evidence of sleep-dependent skill learning has now been demonstrated across a wide variety of skill domains, including the visual (Karni et al., 1994; Gais et al., 2000; Stickgold et al., 2000), auditory (Atienza et al., 2004; Gaab et al., 2004), and motor (Smith and MacNeill, 1994; Fischer et al., 2002; Walker et al., 2002, 2003; Kuriyama et al., 2004) systems. Specifically, sleep has been implicated in the ongoing process of consolidation after initial acquisition, whereby delayed learning could occur in the absence of further practice (Smith, 1995; Stickgold et al., 2001; Walker and Stickgold, 2004).

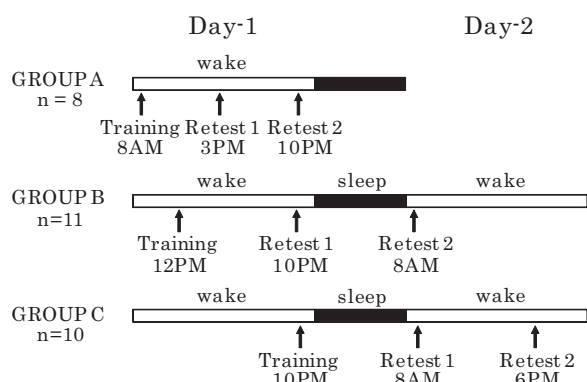
Received May 3, 2008; revised June 5, 2008; accepted Sept. 3, 2008.

This work was supported in part by Research Grant for Nervous and Mental Disorders 15-2, Health Science Grant 17302201 from the Ministry of Health, Labor, and Welfare of Japan, and Grant-in-Aid for Scientific Research 16614018 from the Ministry of Education, Sports, Science, and Culture of Japan.

Correspondence should be addressed to Kenichi Kuriyama, Department of Adult Mental Health, National Institute of Mental Health, National Center of Neurology and Psychiatry, Ogawa-Higashi, Kodaira, Tokyo 187-8502, Japan. E-mail: kenichik@ncnp.go.jp.

DOI:10.1523/JNEUROSCI.2039-08.2008

Copyright © 2008 Society for Neuroscience 0270-6474/08/2810145-06\$15.00/0



**Figure 1.** Study protocol. Twenty-nine subjects were allocated into three experimental groups (A–C). Group A was trained at 8:00 A.M. and retested at 3:00 P.M. and 10:00 P.M. across wakefulness. Group B was trained at 12:00 P.M. (midday) and retested at 10:00 P.M. (after 10 h of wakefulness) and at 8:00 A.M. (10 h later) after a night of sleep. Group C was trained at 10:00 P.M. and retested at 8:00 A.M. (10 h later) after a night of sleep and at 6:00 P.M. (after 10 h of wakefulness).

We hypothesized that improvement in WM performance as measured by the *n*-back procedure could be facilitated by post-training physiological sleep similar to that observed in other skill domains. In this study, we made a particular attempt to discriminate the possible effects of time elapsed after training, posttraining brain state (sleep or wakefulness), and circadian fluctuations in the improvement of *n*-back task performance.

## Materials and Methods

### Participants

A total of 29 right-handed healthy subjects (mean age, 21.9 years; range, 19–26 years; 19 females) were randomly assigned to three different groups (described below). Subjects had no previous history of drug or alcohol abuse or of neurological, psychiatric, or sleep disorders, and were maintaining a constant sleep schedule. They were instructed to be drug-, alcohol-, and caffeine-free for 24 h before, and during, the study period. All procedures for the study were in accordance with the guidelines outlined in the Declaration of Helsinki. The study protocol was approved by the Intramural Research Board of the National Center of Neurology and Psychiatry, and all subjects provided written informed consent to participate in the study.

### Working memory task

We used a spatial variant of the *n*-back WM task, which has been widely used to measure spatial WM with a sustained attention component (Gevins and Cuttillo, 1993; Callicott et al., 1998, 1999; McEvoy et al., 1998), for all three experimental groups (groups A–C) (for details, see Fig. 1 and below). Subjects performed the *n*-back WM task with nine increasing levels of difficulty ( $n = 1–9$ ), using a standard PC. Four large dots presented in a single row were displayed on the screen, indicating the four possible places where a stimulus could appear (Fig. 2). The stimulus consisted of one dot changing color. Subjects were instructed to respond by pushing one of four buttons on a response button box with the right fingers as quickly and as accurately as possible when the next stimulus appeared. The layout of the four buttons corresponded spatially to the four possible positions in which the stimulus appeared. Responses were to be made after a delay of *n* (load level) in *n*-back stimuli. The load level was shown before stimulation began throughout the entire experimental task. The different load levels were run in blocks of  $20 + n$  stimuli; thus, 20 responses were obtained at each load level. The interstimulus interval was set at 500 ms, and each stimulus was displayed for 1500 ms; each block lasted a total of 41,500–58,500 ms. At each level, subjects performed three trials separated by 15,000 ms rest periods, with the scores being averaged at the end of the three trials. The stimuli were set in randomized order for each test session. Subjects completed all load levels

( $n = 1–9$ ) three times in each test session (see below, Experimental design).

### Performance measures

Performance was evaluated by using both the average percentage of correct responses (accuracy) and the average RT at each different load level. These provided measures of improvement in the throughput and the processing speed of the WM, respectively. The detection threshold for a given session was defined as the maximum *n*-back accuracy level (NL) at which the subject's accuracy exceeded at least 80%. RT was also calculated for each session.

### Experimental design

The 29 subjects were assigned to the three experimental groups listed below, and each group underwent a specific schedule consisting of an initial training session and two retest sessions. Subjects performed a spatial *n*-back WM task ( $n = 1–9$ ) in each test session. All morning retests were performed at least 1 h after awakening. Just before the initial training session, each subject performed the spatial *n*-back task ( $n = 0–4$ ) to become familiar with the PC procedure. Retest schedules (Fig. 1) were as follows.

**Group A: continued WM task training across wakefulness.** To determine whether the simple passage of time (across wakefulness) led to improvement in WM performance, eight subjects (mean age, 22.3 years; range, 19–26 years; 5 females) were retested at 7 h intervals across the day after initial training at 8:00 A.M. (i.e., retests at 3:00 P.M. and 10:00 P.M.)

**Group B: continued WM task training followed by 10 h wakefulness and then sleep.** To determine whether subsequent sleep showed any marked improvement in WM performance over wakefulness, 11 subjects (mean age, 21.4 years; range, 19–24 years; 7 females) were trained at 12:00 P.M. (midday) and retested once at 10:00 P.M. after 10 h of wakefulness, and then again at 8:00 A.M. the next morning after a night of sleep.

**Group C: continued WM task training followed by the immediate 8 h sleep and then wakefulness.** To determine whether the improvement of WM performance required sustained wakefulness just after the initial training, 10 subjects (mean age, 20.2 years; range, 19–22 years; 7 females) were trained at 10:00 P.M. followed by an immediate 8 h sleep and then retested at 8:00 A.M. the next morning, and again later at 6:00 P.M. on the same day.

At each training and retesting point, all subjects performed a simple reaction task, which provided their simple reaction time (SRT), a standard measure of subjective alertness (Lorenzo et al., 1995; Corsi-Cabrera et al., 1996). The amount of overnight sleep for each subject in each experimental group was estimated using a self-recorded sleep log.

### Statistics

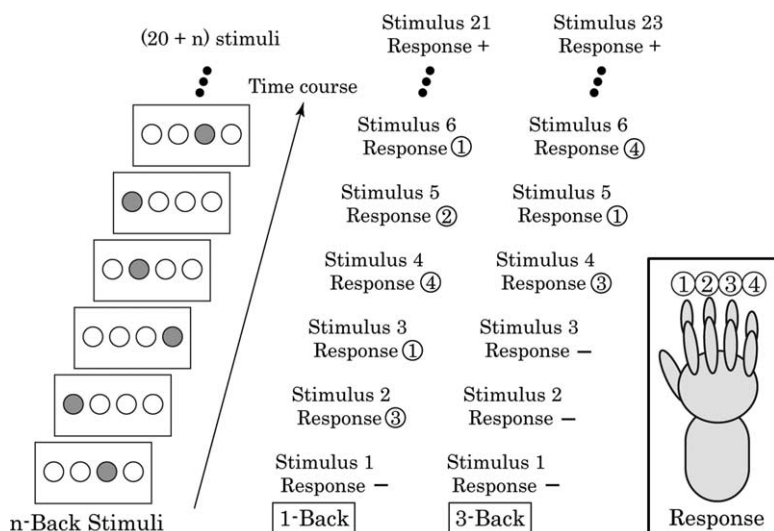
Two-way factorial ANOVA was applied to detect the group and test-session differences in SRT performance. One-way factorial ANOVA was applied to compare the amount of sleep in the previous night of the experiment among three groups. The  $\chi^2$  test was used to compare gender distribution of the study subjects among the three groups.

Two-way factorial ANOVA was applied to detect the load level and gender differences in baseline *n*-back task performance in the three experimental groups (A–C), as well as to compare the improvement of *n*-back task performance among the three groups by 3 (experimental groups)  $\times$  3 (test sessions) and by 3 (experimental groups)  $\times$  2 (inter-sessions; retest 1 minus initial training vs retest 2 minus retest 1) comparison. After the analyses, we used one-way factorial ANOVA to detect the possible role of posttraining sleep in the improvement of *n*-back task performance. All ANOVA were followed by Bonferroni's *post hoc* test. Results are shown as mean and SEM values. A *p* value of  $<0.05$  ( $<0.0167$  in Bonferroni's *post hoc* analysis) was considered to indicate significance.

## Results

### Sleep quality and alertness

Two-way ANOVA revealed no significant differences in SRT within the three experimental groups ( $F_{(2,78)} = 1.920$ ;  $p = 0.1535$ ;  $507.5 \pm 18.0$  vs  $468.7 \pm 12.3$  vs  $486.1 \pm 10.4$  ms for groups A–C, respectively) or within the three test sessions ( $F_{(2,78)} = 0.076$ ;  $p =$



**Figure 2.** The spatial variant of the *n*-back working memory task configurations. Four large white dots aligned in a single row were displayed on the screen, indicating the four possible places where a stimulus could appear. The stimulus consisted of one of the four dots changing from white to red. Subjects had to respond to the stimulus by pushing one of four buttons, which were arranged in lines corresponding spatially to the four possible stimulus positions, with the right fingers as quickly and accurately as possible when the next stimulus appeared. Responses had to be made after a delay of *n* (*n*-back) stimuli. The different load levels (1–9) were run in blocks of  $20 + n$  stimuli; thus, 20 responses were obtained at each load level. The interstimulus interval was set at 500 ms, and a stimulus was displayed for 1500 ms. Arrangement of the stimuli was randomized in each test session.

0.9267;  $488.6 \pm 17.2$  vs  $482.5 \pm 11.5$  vs  $485.1 \pm 11.4$  ms for initial training, retest 1, and retest 2, respectively). In addition, no interaction in SRT between the experimental groups and test sessions ( $F_{(4,78)} = 0.250$ ;  $p = 0.9090$ ) was found. One-way ANOVA detected no significant difference in the amount of sleep among the experimental groups ( $F_{(2,26)} = 2.872$ ;  $p = 0.0747$ ;  $7.31 \pm 0.19$  vs  $7.55 \pm 0.22$  vs  $6.85 \pm 0.22$  h for groups A–C, respectively). These findings indicate that there were no clear differences in vigilance level among the subjects in each test session.

### Initial training analyses

We analyzed the difficulty profiles of accuracy and RT for each load level in the *n*-back task for each experimental group (A–C). Two-way ANOVA revealed significant differences in accuracy among load levels ( $F_{(8,294)} = 114.1$ ;  $p < 0.001$ ) (Fig. 3), but not among experimental groups ( $F_{(2,294)} = 2.905$ ;  $p = 0.0567$ ). No interaction was seen between load levels and experimental groups ( $F_{(16,294)} = 0.954$ ;  $p = 0.5086$ ) in terms of accuracy. A *post hoc* test for the load level revealed a significant decrement in accuracy between load level 5 and load level 6 ( $p < 0.0001$ ), in that task difficulty gradually increased with an increase in trial number (*n*) up to 5; it then sharply increased at trial 6 and thereafter remained high.

Two-way ANOVA revealed no significant differences in RT either among load levels ( $F_{(8,294)} = 0.304$ ;  $p = 0.9641$ ) (Fig. 3) or among experimental groups ( $F_{(2,294)} = 2.520$ ;  $p = 0.0826$ ). No interaction was seen between load levels and experimental groups ( $F_{(16,294)} = 0.122$ ;  $p > 0.9999$ ) in terms of RT.

Gender effects on WM performance have been speculated in a previous study (Duff and Hampson, 2001). We therefore examined gender distribution in each experimental group. The  $\chi^2$  test revealed no significant gender distribution among the experimental groups ( $\chi^2 = 0.138$ ;  $p = 0.9331$ ), suggesting that each group included almost equal gender distribution. Two-way ANOVA revealed no significant gender difference in accuracy (male vs female;  $68.34 \pm 2.39\%$  vs  $67.18 \pm 1.84\%$ ;  $F_{(1,243)} =$

$0.696$ ;  $p = 0.4048$ ), but there were significant load level differences in accuracy ( $F_{(8,243)} = 108.3$ ;  $p < 0.001$ ) on the WM task, although no significant interaction was observed between gender and load level in terms of accuracy ( $F_{(8,243)} = 0.908$ ;  $p = 0.5106$ ). Likewise, two-way ANOVA revealed no significant gender difference in RT (male vs female;  $299.5 \pm 13.4$  ms vs  $325.7 \pm 13.0$  ms;  $F_{(1,243)} = 1.580$ ;  $p = 0.2100$ ) and no significant load level difference ( $F_{(8,243)} = 0.245$ ;  $p = 0.9817$ ) in RT. Moreover, no significant interaction was seen between gender and load level in terms of RT ( $F_{(8,243)} = 0.294$ ;  $p = 0.9842$ ). These findings indicate that there was no gender difference in initial training performance.

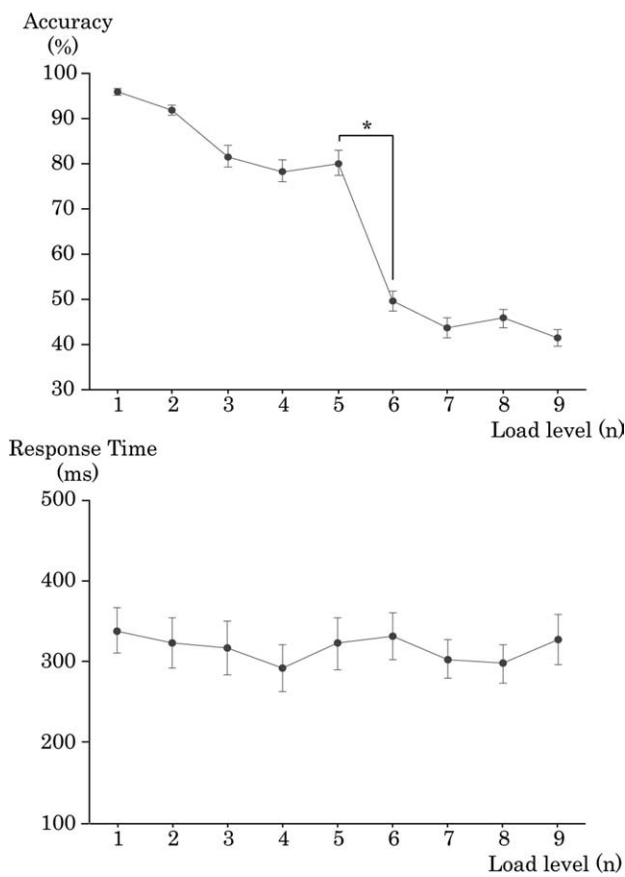
### Analyses for experimental group $\times$ test-session interaction

Two-way ANOVA (3 experimental groups  $\times$  3 test sessions) showed significant group and test-session effects on NL ( $F_{(2,78)} = 4.147$ ,  $p = 0.0194$ ;  $F_{(2,78)} = 7.019$ ,  $p = 0.0016$ ; respectively) (Fig. 4), but no significant group  $\times$  test-session interaction ( $F_{(4,78)} = 1.453$ ;  $p = 0.2246$ ). Two-way ANOVA (3 groups  $\times$  2 intersessions) showed significant group effects on NL improvement ( $F_{(2,52)} = 3.686$ ;  $p = 0.0318$ ) and a significant group  $\times$  intersession interaction ( $F_{(2,52)} = 15.857$ ;  $p < 0.0001$ ) (Fig. 5), but no significant intersession effect on NL improvement ( $F_{(1,52)} = 0.248$ ;  $p = 0.6205$ ). A *post hoc* test for NL improvement revealed a significant difference between groups A and C ( $p = 0.0109$ ), and a trend toward intergroup differences between groups A and B ( $p = 0.0491$ ). These findings suggest that the three experimental groups showed different time profiles of NL improvement; NL improvement during posttraining sleep was significantly greater than that during wakefulness, and the acquired NL improvement seemed to be maintained for at least 10 h after posttraining sleep. As a result, average NL improvement in subjects who experienced posttraining sleep (groups B and C) was greater than that in subjects who went through the same period of wakefulness (group A) (Fig. 5).

Two-way ANOVA (3 experimental groups  $\times$  3 test sessions) showed neither significant group nor test-session effects on RT ( $F_{(2,78)} = 0.719$ ,  $p = 0.4904$ ;  $F_{(2,78)} = 0.027$ ,  $p = 0.9738$ ; respectively) (Fig. 4); moreover, no significant group  $\times$  test-session interaction was observed ( $F_{(4,78)} = 0.307$ ;  $p = 0.8723$ ).

### Group A: continued WM task training across wakefulness

One-way ANOVA revealed no significant test-session difference in NL in group A subjects ( $F_{(2,21)} = 0.218$ ;  $p = 0.8058$ ). Compared with the NL for the initial training ( $3.13 \pm 0.72$ ), we observed a subtle but not statistically significant improvement in NL at 3:00 P.M. ( $3.63 \pm 0.71$ , by 16.0% vs initial training) and at 10:00 P.M. ( $3.75 \pm 0.70$ , by 19.8% vs initial training) (Fig. 4A), suggesting that the simple passage of time across wakefulness produced no significant improvement in WM performance beyond that expected on the basis of continued rehearsal.



**Figure 3.** Initial training performances in all groups. Accuracy (top) and RT (bottom) in the initial training session for each load level are plotted. Filled circles with error bars represent mean and SEM values in each panel. Significant interload level difference in accuracy was observed between level 5 and level 6; the accuracy linearly decreased as the task difficulty increased up to level 5 before rapidly dropping at level 6 and remaining low at <50% thereafter. We observed no significant interload level difference in RT. \* $p < 0.0001$ .

### Group B: continued WM task training after wakefulness and then sleep

One-way ANOVA revealed significant test-session differences in NL in group B subjects ( $F_{(2,30)} = 5.104$ ;  $p = 0.0124$ ). A *post hoc* test for NL revealed significant differences between the following test sessions: initial training versus retest 2 ( $4.00 \pm 0.40$  vs  $5.73 \pm 0.49$ ;  $p = 0.0081$ ) and retest 1 versus retest 2 ( $4.09 \pm 0.39$  vs  $5.73 \pm 0.49$ ;  $p = 0.0116$ ).

Similarly to group A subjects, group B subjects demonstrated no significant increase in NL at 10:00 P.M. (by 2.25% vs initial training;  $p = 0.8822$ ) (Fig. 4B), but demonstrated a significant increase in NL at retest 2 the next morning (by 40.1% vs retest 1 before sleep). These data suggest that the significant improvement in NL was obtained not during the 10 h of wakefulness just after initial training but after the posttraining sleep 10 h or more after the initial training.

### Group C: continued WM task training after sleep and then wakefulness

One-way ANOVA revealed significant test-session differences in NL in group C subjects ( $F_{(2,27)} = 14.678$ ;  $p < 0.0001$ ). A *post hoc* test for NL revealed significant differences between the following test sessions: initial training versus retest 1 ( $3.30 \pm 0.35$  vs  $4.90 \pm 0.20$ ;  $p = 0.0002$ ) and initial training versus retest 2 ( $3.30 \pm 0.35$  vs  $5.10 \pm 0.28$ ;  $p < 0.0001$ ).

After a night of posttraining sleep, a significant increase in NL

was apparent at 8:00 A.M. the next morning compared with the initial training scores (by 48.5%) (Fig. 4C). However, an additional 10 h of wakefulness produced no significant change in NL compared with the retest 1 scores (by 4.08%;  $p = 0.4577$ ).

## Discussion

### Sleep-dependent facilitation of WM performance improvement

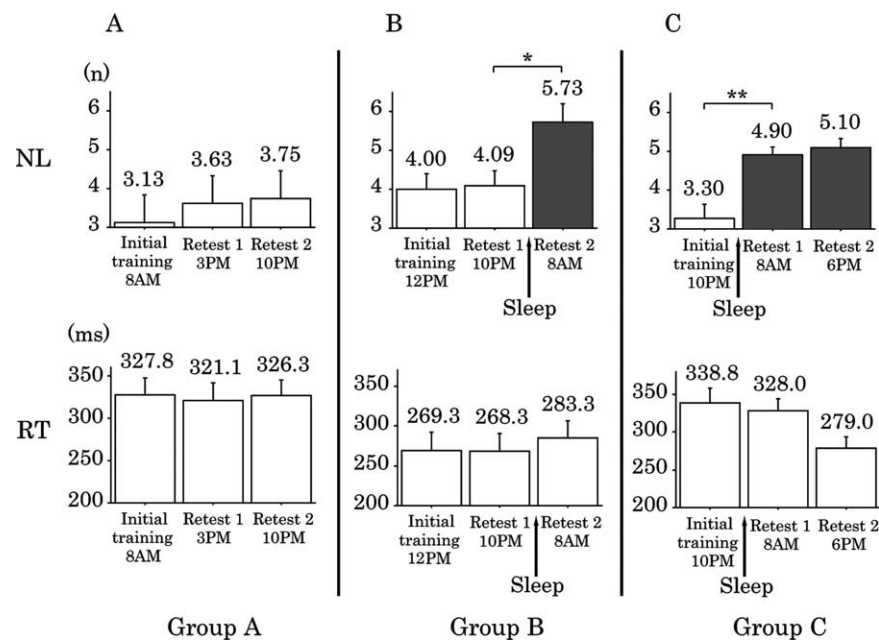
Although the NLs at the initial training session were comparable among all experimental groups or by gender, subjects demonstrated remarkably different time courses of subsequent NL improvements that were specifically dependent on the timing of posttraining sleep. Subjects trained at 12:00 P.M. (midday) demonstrated no significant improvement when retested after 10 h of wakefulness, but showed a significant improvement at 8:00 A.M. after a night of posttraining sleep (by 40.1% in group B) (Fig. 4). Similarly, subjects trained at 10:00 P.M. showed a significant overnight improvement (by 48.5% in group C) (Fig. 4), but no significant additional improvement during a further 10 h of wakefulness. Thus, significant improvements were acquired only across a night of posttraining sleep and not over a similar period of wakefulness, regardless of whether the time awake or time asleep came first.

The possibility that circadian factors confounded the learning profiles after 10 h of wakefulness is unlikely. The initial training session was similar for subjects trained at 8:00 A.M., 12:00 P.M., or 10:00 P.M., as was the case for objective ratings of alertness across all testing points. Thus, we consider sleep itself to be the most likely source of the improvement in NL on the *n*-back task.

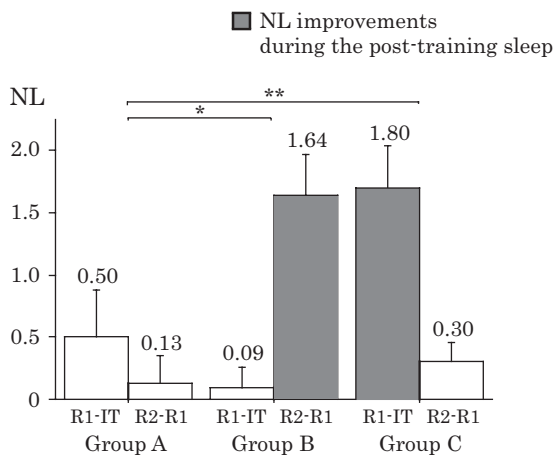
RT has been considered to be a good indicator of skill performance improvement (Fischer et al., 2002; Walker et al., 2002; Kuriyama et al., 2004), but in the present study, it rarely seemed to reflect improvement in WM task performance. In our subjects, RT values varied independently of the comparative difficulty of the *n*-back task, which is in contrast to the values observed in the accuracy profiles for initial training. A previous study involving repetitive WM task training has also shown marked improvement in accuracy values over a 1–2 d period, although RT values improved slowly over a 4–5 d period (Olesen et al., 2004). Taking together, these findings suggest that RT may reflect different levels of improvement in WM task performance on the basis of accuracy values. WM performance is considered to be a result of plural cognitive processing (Gevins and Cuttolo, 1993; Owen et al., 1996). The RT value possibly reflects the total skill performance of WM, whereas accuracy reflects the capacity limitation of WM.

### Possible contribution of improvement in *n*-back accuracy to generalized improvement of WM performance

The results of recent investigations using a spatial variant of the WM task to examine the feasible number of items for both storing temporary and manipulating data converged on around three or four items (Luck and Vogel, 1997; Cowan, 2001; Saults and Cowan, 2007). The accuracy index for the *n*-back task has been established as a reflection of WM capacity in previous studies, and has been used for investigating individual or age variation in WM capacity (Oberauer, 2005; Mattay et al., 2006). Consistent with these previous findings, our subjects in all experimental groups showed NL around 3 or 4 at the initial training session. In the postsleep session, the NL for groups B and C increased up to around 5 and 6. Thus, the NL improvement was acquired across sleep, suggesting that dynamic changes in the WM process were executed during sleep, as has been observed in other cases of skill



**Figure 4.** Time courses of improvement in *n*-back task performance. Time courses of NL (top) and RT (bottom) are displayed. **A–C**, Bars and error bars represent mean and SEM values in groups A–C, respectively. In the sessions after posttraining sleep, we observed significant improvement in NL compared with those before sleep (groups B and C; black bars), but RT showed no significant benefit from sleep. \* $p < 0.0167$ ; \*\* $p < 0.001$ .



**Figure 5.** Intersession differences in improvement in *n*-back accuracy. Bars and error bars represent mean and SEM values of intersession differences in NL in each group. Left and right bars in each experimental group show intersession differences in NL between initial training (IT) and retest 1 (R1) sessions, and between R1 and R2 sessions, respectively. Filled bars represent NL improvements during the posttraining sleep in groups B and C. *Post hoc* test revealed a significant intergroup difference in NL between groups A and C (\*\* $p = 0.0109$ ) and a trend toward intergroup difference in NL between groups A and B (\* $p = 0.0491$ ).

learning (Stickgold et al., 2000; Walker et al., 2002; Kuriyama et al., 2004).

Some studies focusing on sleep-dependent skill learning have emphasized that the sleep-dependent gains seen in procedural skills were specific to the stimulus materials used, which therefore could not affect skill performance using other stimuli (Korman et al., 2003; Walker et al., 2003). Thus, the sleep-dependent benefit on the *n*-back task observed in the present study might be limited to the particular stimulus we used and may not be generalizable to other WM tasks.

However, Jaeggi et al. (2008) have recently demonstrated the

landmark finding that repetitive training on a spatial *n*-back task improved not only spatial *n*-back performance but also auditory *n*-back performance simultaneously. Moreover, they found that the performance improvement on the *n*-back task could involve the improvement of general fluid intelligence as measured by a standardized fluid intelligence test (Jaeggi et al., 2008). Olesen et al. (2004) reported training-induced changes in cortical activity after 5 weeks of WM training, and the increment of WM load showed increased cortical activity in the middle frontal gyrus and superior and inferior parietal cortices, where activity changes are known to be less specific to various stimuli that drive cognitive performance (Klingberg, 1998; Duncan and Owen, 2000).

Together, these findings suggest that improvement in WM performance might not depend on the type of stimulus used, and that the sleep-dependent improvement in WM performance seen in the present study may lead to various improvements in WM performance, and furthermore, in the general capacity of WM.

WM capacity is an important factor in a wide range of cognitive abilities, including general fluid intelligence (Conway et al., 2003; Colom et al., 2007), and our finding together with that of Jaeggi et al. (2008) suggests that posttraining sleep with appropriate timing could be a potent facilitating factor in WM training, leading to the advancement of individual general fluid intelligence.

## References

- Atienza M, Cantero JL, Stickgold R (2004) Posttraining sleep enhances automaticity in perceptual discrimination. *J Cogn Neurosci* 16:53–64.
- Baddeley AD, Hitch GJ (1974) Working memory. In: *The psychology of learning and motivation* (Bower GA, ed.), pp 47–89. New York: Academic.
- Callicott JH, Ramsey NF, Tallent K, Bertolino A, Knable MB, Coppola R, Goldberg T, van Gelderen P, Mattay VS, Frank JA, Moonen CT, Weinberger DR (1998) Functional magnetic resonance imaging brain mapping in psychiatry: methodological issues illustrated in a study of working memory in schizophrenia. *Neuropsychopharmacology* 18:186–196.
- Callicott JH, Mattay VS, Bertolino A, Finn K, Coppola R, Frank JA, Goldberg TE, Weinberger DR (1999) Physiological characteristics of capacity constraints in working memory as revealed by functional MRI. *Cereb Cortex* 9:20–26.
- Colom R, Jung RE, Haier RJ (2007) General intelligence and memory span: evidence for a common neuroanatomic framework. *Cogn Neuropsychol* 24:867–878.
- Conway AR, Kane MJ, Engle RW (2003) Working memory capacity and its relation to general intelligence. *Trends Cogn Sci* 7:547–552.
- Corsi-Cabrera M, Arce C, Ramos J, Lorenzo I, Guevara MA (1996) Time course of reaction time and EEG while performing a vigilance task during total sleep deprivation. *Sleep* 19:563–569.
- Cowan N (2001) The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci* 24:87–114; discussion 114–185.
- Duff SJ, Hampson E (2001) A sex difference on a novel spatial working memory task in humans. *Brain Cogn* 47:470–493.
- Duncan J, Owen AM (2000) Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci* 23:475–483.
- 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.

- Fry AF, Hale S (1996) Processing speed, working memory and fluid intelligence: evidence for a developmental cascade. *Psychol Sci* 7:237–241.
- Gaab N, Paetzold M, Becker M, Walker MP, Schlaug G (2004) The influence of sleep on auditory learning: a behavioral study. *Neuroreport* 15:731–734.
- Gais S, Plihal W, Wagner U, Born J (2000) Early sleep triggers memory for early visual discrimination skills. *Nat Neurosci* 3:1335–1339.
- Gevins A, Cuttillo B (1993) Spatiotemporal dynamics of component processes in human working memory. *Electroencephalogr Clin Neurophysiol* 87:128–143.
- Goldman-Rakic PS (1996) The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive. *Philos Trans R Soc Lond B Biol Sci* 51:1445–1453.
- Hale S, Bronik MD, Fry AF (1997) Verbal and spatial working memory in school-age children: developmental differences in susceptibility to interference. *Dev Psychol* 33:364–371.
- Jaeggi SM, Buschkuhl M, Jonides J, Perrig WJ (2008) Improving fluid intelligence with training on working memory. *Proc Natl Acad Sci U S A* 105:6829–6833.
- Jansma JM, Ramsey NF, van der Wee NJ, Kahn RS (2004) Working memory capacity in schizophrenia: a parametric fMRI study. *Schizophr Res* 68:159–171.
- Karni A, Tanne D, Rubenstein BS, Askenasy JJ, Sagi D (1994) Dependence on REM sleep of overnight improvement of a perceptual skill. *Science* 265:679–682.
- Klingberg T (1998) Concurrent performance of two working memory tasks: potential mechanisms of interference. *Cereb Cortex* 8:593–601.
- 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.
- Kuriyama K, Stickgold R, Walker MP (2004) Sleep-dependent learning and motor-skill complexity. *Learn Mem* 11:705–713.
- Kyllonen PC, Christal RE (1990) Reasoning ability is (little more than) working memory capacity?! *Intelligence* 14:389–433.
- Lorenzo I, Ramos J, Arce C, Guevara MA, Corsi-Cabrera M (1995) Effect of total sleep deprivation on reaction time and waking EEG activity in man. *Sleep* 18:346–354.
- Luck SJ, Vogel EK (1997) The capacity of visual working memory for features and conjunction. *Nature* 390:279–281.
- Mattay VS, Fera F, Tessitore A, Hariri AR, Berman KF, Das S, Meyer-Lindenberg A, Goldberg TE, Callicott JH, Weinberger DR (2006) Neurophysiological correlates of age-related changes in working memory capacity. *Neurosci Lett* 392:32–37.
- McEvoy LK, Smith ME, Gevins A (1998) Dynamic cortical networks of verbal and spatial working memory: effects of memory load and task practice. *Cereb Cortex* 8:563–574.
- McNab F, Klingberg T (2008) Prefrontal cortex and basal ganglia control access to working memory. *Nat Neurosci* 11:103–107.
- Miller GA (1956) The magical number seven, plus minus two; some limits on our capacity for processing information. *Psychol Rev* 63:81–97.
- Oberauer K (2005) Binding and inhibition in working memory: individual and age differences in short-term recognition. *J Exp Psychol Gen* 134:368–387.
- Olesen PJ, Westerberg H, Klingberg T (2004) Increased prefrontal and parietal activity after training of working memory. *Nat Neurosci* 7:75–79.
- Owen AM, Evans AC, Petrides M (1996) Evidence for a two-stage model of spatial working memory processing within the lateral frontal cortex: a positron emission tomography study. *Cereb Cortex* 6:31–38.
- Saults JS, Cowan N (2007) A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *J Exp Psychol Gen* 136:663–684.
- Smith C (1995) Sleep states and memory processes. *Behav Brain Res* 69:137–145.
- Smith C, MacNeill C (1994) Impaired motor memory for a pursuit rotor task following Stage 2 sleep loss in college students. *J Sleep Res* 3:206–213.
- Stickgold R, James L, Hobson JA (2000) Visual discrimination learning requires sleep after training. *Nat Neurosci* 3:1237–1238.
- Stickgold R, Hobson JA, Fosse R, Fosse M (2001) Sleep, learning, and dreams: off-line memory reprocessing. *Science* 294:1052–1057.
- Walker MP, Stickgold R (2004) Sleep-dependent learning and memory consolidation. *Neuron* 44:121–133.
- 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.
- Walker MP, Brakefield T, Hobson JA, Stickgold R (2003) Dissociable stages of human memory consolidation and reconsolidation. *Nature* 425:616–620.
- Westerberg H, Klingberg T (2007) Changes in cortical activity after training of working memory: a single-subject analysis. *Physiol Behav* 92:186–192.