HHS Public Access

Author manuscript

Nat Neurosci. Author manuscript; available in PMC 2021 January 20.

Published in final edited form as:

Nat Neurosci. 2020 September ; 23(9): 1150–1156. doi:10.1038/s41593-020-0666-y.

Complementary contributions of NREM and REM sleep to visual learning

Masako Tamaki[1,3], **Zhiyan Wang**[1], **Tyler Barnes-Diana**[1], **DeeAnn Guo**[2], **Aaron V. Berard**[1], **Edward Walsh**[2], **Takeo Watanabe**[1], **Yuka Sasaki**[1,*]

[1]Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, Rhode Island, USA.

[2]Department of Neuroscience, Brown University, Providence, Rhode Island, USA.

[3]Present Address: National Institute of Occupational Safety and Health, Kawasaki, Japan

Abstract

Sleep is beneficial for learning. However, it remains unclear whether learning is facilitated by non-REM (NREM) sleep or by REM sleep, whether it results from plasticity increases or stabilization, and whether facilitation results from learning-specific processing. Here, we trained volunteers on a visual task, and measured the excitatory and inhibitory (E/I) balance in early visual areas during subsequent sleep as an index of plasticity. E/I balance increased during NREM sleep irrespective of whether pre-sleep learning occurred, but it was associated with post-sleep performance gains relative to pre-sleep performance. By contrast, E/I balance decreased during REM sleep but only after pre-sleep training, and the decrease was associated with stabilization of pre-sleep learning. These findings indicate that NREM sleep promotes plasticity, leading to performance gains independent of learning, while REM sleep decreases plasticity to stabilize learning in a learning-specific manner.

Introduction

An accumulating body of evidence demonstrates that sleep is beneficial for various types of learning and memory. However, the underlying neural mechanisms for how sleep impacts learning and memory are poorly understood. At present, there are at least three serious controversies. The first controversy concerns the roles of sleep stages in the facilitation of learning. Sleep is largely categorized into nonrapid eye movement sleep (NREM) sleep and rapid eye movement (REM) sleep[1]. Some research groups have suggested that NREM sleep[2-4] plays an important role in facilitating learning, while other groups emphasized the role of REM sleep[5, 6]. Whether NREM and/or REM sleep contributes to learning is

*Correspondence to: Yuka Sasaki (yuka_sasaki@brown.edu).

Author Contributions

MT, TW and YS designed the research. MT, ZW, TBD, DG, AVB, and EW performed the experiments. MT, EW, ZW and TBD analyzed the data. MT, ZW, TBD, DG, AVB, EW, TW and YS wrote the manuscript.

Competing Interests

The authors declare no competing interests.

unclear[7-9]. The second controversy concerns how sleep is beneficial to learning. At least two different types of learning benefits have been noted: off-line performance gains in learning[6, 10] and resilience to interference[11-13]. Off-line performance gains mean that the learning acquired before sleep is enhanced after sleep without any additional training. The resilience to interference conferred by sleep refers to the decreased amount of retrograde interference, or disruption, with learning acquired before sleep by learning or training on a new task after sleep. Retrograde interference with learning is regarded as a manifestation of the highly plastic learning state that is vulnerable to “rewriting” by subsequent new learning[14]. Elimination of retrograde interference by sleep suggests that learning acquired before sleep is stabilized by sleep. It is unclear how both of these benefits are associated with plasticity states occurring during sleep or how each sleep stage contributes to states of plasticity. The third controversy is whether the facilitation of learning occurs in a learning-specific or learning-independent manner. The former assumes that only synapses or networks that are specifically involved in the acquisition of learning and memory before sleep are changed during subsequent sleep, which leads to performance improvement after sleep[7, 15, 16]. The latter is based on the synaptic homeostasis hypothesis, which is a use-dependent process[4]. In this theory, during wakefulness, synapses that are used are overly strengthened irrespective of learning. During subsequent sleep, these overall synaptic changes are downscaled for synaptic homeostasis[4], and only strong synapses survive.

To resolve these controversial issues, we measured degrees of plasticity and stability during NREM and REM sleep after visual perceptual learning (VPL), defined as improvement in visual task performance after a visual experience[17, 18]. We used human subjects to take advantage of the fact that humans have significantly longer and more clearly distinguished NREM and REM sleep episodes than animals. Although it was regarded as difficult to examine neurochemical processing related to plasticity during sleep in humans, for the first time, we successfully made simultaneous magnetic resonance spectroscopy (MRS) and polysomnography measurements in sleeping humans. Specifically, MRS allowed us to measure the concentrations of glutamate (Glx), an excitatory neurotransmitter, and gamma-aminobutyric acid (GABA), an inhibitory neurotransmitter, in human early visual areas. We calculated the excitation/inhibition (E/I) balance, defined as the concentration of Glx divided by the concentration of GABA, because the E/I balance is regarded as a reliable index of the degree of plasticity and stability in early visual areas[19-21]. The visual critical period in rodents[19] starts with a relatively high E/I balance in early visual areas and ends with a lower E/I balance. In human studies, a higher E/I balance is proportional to the degree of plasticity, as measured by psychophysics, and negatively correlated with the degree of stability of VPL in early visual areas[20, 21].

By measuring the E/I balance during sleep, we found complementary neurochemical processing during NREM and REM sleep that was highly associated with the plasticity and stability of VPL in human brains. First, the plasticity of early visual areas increased during NREM sleep, as shown by an increased E/I balance, which predicted off-line performance gains well. Importantly, the E/I balance during NREM sleep increased irrespective of whether learning occurred before sleep, indicating that plasticity increases in early visual areas during NREM sleep in a learning-independent manner. Second, REM sleep following NREM sleep played a role in stabilization, making learning before sleep resilient to

retrograde interference by new learning after sleep to protect performance gains developed during NREM sleep. Without REM sleep, the performance gains that would have occurred by NREM sleep were nullified by new learning after sleep. During REM sleep, the E/I balance, which increased during NREM sleep, decreased to a lower level than that measured at baseline during wakefulness. The decrease in E/I balance during REM sleep was correlated with the degree of resilience to retrograde interference in VPL. Importantly, the decrease in E/I balance during REM sleep was not observed without presleep learning and therefore was specific to learning.

These results consistently suggest that NREM sleep and REM sleep play distinctive roles that are subserved by the opposing directions of functional and neurochemical processing. Thus, each of the opponent concepts, such as performance gains vs. stabilization, the contribution of NREM sleep vs. REM sleep, and learning independence vs. learning dependence, may not be issues of controversy but properties of one of two distinctive functions.

Results

Experiment 1

In Experiment 1, we examined the roles of NREM and REM sleep in sleep facilitation of performance gains and stabilization of presleep learning. To this end, we tested how the E/I balance (Glx/GABA), which is regarded as an index of plasticity in early visual areas, changes across NREM and REM sleep and how it is related to off-line performance gains and stabilization of presleep learning. In particular, we tested the complementary processing hypothesis that off-line performance gains and stabilization of learning are subserved by complementary neurochemical processing during NREM and REM sleep, respectively.

This hypothesis was inspired by the results of studies of VPL during wakefulness. Retrograde interference was found during wakefulness if the interval between two trainings was within a few hours[22, 23]. Such interference is attributed to the incomplete stabilization during wakefulness of the first learning: soon after training for the first task, the state of the learning is still highly plastic, and the passage of a few hours is necessary to stabilize the learning[14]. In support of this, the E/I balance increased after training and decreased to the baseline value as stabilization proceeded[20, 21]. These results of previous studies during wakefulness lead to the current hypothesis: if plasticity increases during sleep for presleep learning, this should be followed by stabilization, as shown by the decreased plasticity for presleep learning to survive retrograde interference by new learning after sleep.

To actualize such plasticity increases followed by stabilization during sleep, NREM sleep and REM sleep that follows NREM sleep may be suitable. If this hypothesis is correct, the following complementary processing should be revealed: during NREM sleep after training on a visual task, the E/I balance in early visual areas should increase with off-line performance gains, whereas during REM sleep, the E/I balance should decrease with the resilience of retrograde interference. Moreover, if REM sleep is necessary to stabilize off-line performance gains, off-line performance gains should not be observed in subjects who had only NREM sleep, without REM sleep in the face of retrograde interference.

In Experiment 1, we tested whether sleep stabilizes off-line performance gains in presleep learning. We trained subjects (see Supplementary Table 1 for more detailed information about subjects) on two different texture discrimination tasks (TDTs[24]) before and after sleep that lasted approximately 90 min (see Fig. 1a and Methods for more details). We termed the presleep TDT Task-A and the postsleep TDT Task-B. Importantly, the orientations of the background elements in Task-A and Task-B were orthogonal to each other, as sequential trainings on TDTs with orthogonal background orientations are known to cause interference in learning[25], even when the trained visual field where targets are presented is consistent across the sequential trainings. We used TDTs for Task-A and Task-B in this interference paradigm to examine whether off-line performance gains in Task-A were stabilized during sleep and resilient to retrograde interference from Task-B. The performance, which was the threshold stimulus-to-mask onset asynchrony (SOA, see Texture discrimination task (TDT) in Methods) at which subjects showed 80% of responses correct for Task-A, was measured in each of 4 test sessions (Fig. 1b, see Methods).

The sleep session took place inside the MRI scanner with polysomnography (see Methods for more details and Supplementary Table 2 for sleep structures). Glx and GABA concentrations were measured from early visual areas (see Extended Data Fig. 1 for an example location) using MRS, and the mean E/I balance was calculated for NREM sleep and REM sleep relative to baseline, which was measured during wakefulness. See Co-registration of MRS data and sleep stages in Methods for more details and Extended Data Fig. 2 for example spectra from MRS.

To specifically examine the role of REM sleep, the subjects were classified into two groups (see Experimental Designs in Methods for more details). The NREM+REM group consisted of subjects who showed both NREM and REM sleep (n = 10 subjects) during the sleep session, whereas the NREM-only group consisted of subjects who showed NREM sleep without REM sleep (n = 9 subjects).

The off-line performance gains in Task-A were defined by [(the performance change between the second and third test sessions)/(the performance in the second test session)] x 100% (see Fig. 1b). Significant off-line performance gains were observed in all subjects (one sample t-test, n = 19 subjects, $t_{18} = 6.40$, $p < 0.001$, Cohen's $d = 1.47$, 95% confidence interval (CI) [10.45, 20.67]).

Next, we examined the relationship between the degree of off-line performance gain and E/I balance during NREM sleep and REM sleep. First, compared to that at baseline, the E/I balance increased during NREM sleep (Fig. 1c, one sample t-test, n = 19 subjects, $t_{18} = 4.77$, $p < 0.001$, Cohen's $d = 1.09$, 95% CI [4.61, 11.88]) and decreased during REM sleep (one sample t-test, n = 10 subjects, $t_9 = 3.21$, $p = 0.011$, Cohen's $d = 1.02$, 95% CI [−13.82, −2.40], for the NREM+REM group only, as the NREM-only group did not have REM sleep). Second, off-line performance gain was significantly correlated with E/I balance during NREM sleep (n = 19 subjects, Pearson's $r_{17} = 0.62$, $p = 0.005$, 95% CI [0.23, 0.84], with a Bonferroni adjusted alpha level of 0.025 (0.05/2)), as shown in Fig. 1d, whereas it was not significantly correlated with E/I balance during REM sleep (n = 10 subjects, Pearson's $r_8 = 0.33$, $p = 0.353$, for the NREM+REM group only). Third, for further

confirmation of the second point, we performed a partial correlation analysis. We obtained a robust and positively significant partial correlation between off-line performance gain and E/I balance during NREM sleep while controlling for the E/I balance during REM sleep (n = 10 subjects, Pearson's partial correlation $r_7 = 0.75$, $p = 0.020$, 95% CI [0.23, 0.94], for the NREM+REM group only). These results demonstrate that the E/I balance during NREM sleep, but not the E/I balance during REM sleep, plays an important role in off-line performance gain.

Then, we examined whether stabilization of presleep learning occurred during sleep. If performance on Task-A does not decrease after training on Task-B, this should indicate resilience to interference from Task-B. Conversely, if the performance on Task-A decreases after training on Task-B, this should indicate retrograde interference from Task-B. Here, we calculated the degree of the resilience of presleep learning, which was shown as the performance change on Task-A with the following equation: [(the performance change between the third and fourth test sessions)/(the performance in the third test session)] x 100% (see Fig. 1b). Here, we specifically examined whether REM sleep was involved in the stabilization of learning.

First, we found that the degree of resilience to retrograde interference was significantly larger for the NREM+REM group than for the NREM only group (Fig. 1e, n = 19 subjects, independent-samples *t*-test, $t_{17} = 4.58$, $p < 0.001$, Cohen's $d = 2.10$, 95% CI [5.58, 13.50]). Second, we found that the E/I balance, which was below zero in 9 out of 10 subjects during REM sleep, was also significantly negatively correlated with the degree of resilience in the NREM+REM group (Fig. 1f, n = 10 subjects, Pearson's $r_8 = -0.73$, $p = 0.016$, 95% CI [−0.93, −0.19]). Third, there was a robust negative partial correlation between resilience to retrograde interference and E/I balance during REM sleep while controlling for the E/I balance during NREM sleep (n = 10 subjects, Pearson's partial correlation $r_7 = -0.74$, $p = 0.024$, 95% CI [−0.93, −0.20] for the NREM+REM group only, as the NREM-only group did not have REM sleep). Fourth, in the NREM-only group, the performance of Task-A increased immediately after the sleep session in the third test session in most subjects (gray circles in Fig. 1d; n = 9 subjects, one sample *t*-test, $t_8 = 5.63$, $p < 0.001$, $d = 1.88$, 95% CI [10.89, 26.02]) but significantly decreased after training on Task-B in the fourth test session, showing clear retrograde interference (gray bar, Fig. 1e, one sample *t*-test, n = 9 subjects, $t_8 = 3.60$, $p = 0.007$, $d = 1.20$, 95% CI [−22.74, −4.98]). This result suggests that resilience to retrograde interference requires REM sleep. Fifth, sleepiness (Supplementary Tables 3-6) and the initial performance levels (Supplementary Table 7) were not significantly different between the two groups. Therefore, they are unlikely to have caused the difference in resilience to retrograde interference between the two groups. Sixth, none of the E/I balance changes or performance changes were associated with the duration of sleep (Supplementary Tables 8-10). In addition, we did not find a significant change in the frequency drift between the first and last MRS runs (Supplementary Table 11). These results suggest that the duration of sleep or MRS data quality is not a major factor impacting the E/I balance or performance changes.

These results consistently suggest two distinctive mechanisms as follows: during NREM sleep, plasticity increases, which leads to off-line performance gains, whereas during REM

sleep, plasticity decreases to less than that during wakefulness, which eliminates or reduces retrograde interference.

Experiment 2

We also conducted a behavioral experiment (Experiment 2) with a larger number of subjects (n = 38 subjects; Supplementary Table 1 for subjects' information) to test whether we could replicate the same behavioral results as in Experiment 1. The results were consistent with those in Experiment 1 (Extended Data Fig. 3). The subjects who had both NREM sleep and REM sleep showed resilience of Task-A to Task-B, whereas the subjects who had only NREM sleep showed interference from Task-B on Task-A.

Experiment 3

Do increases and decreases in E/I balances in early visual areas during NREM sleep and REM sleep occur only after learning or are these changes innate with sleep itself? If the E/I balance changes during sleep stages are not learning-dependent, without presleep learning, the E/I balance should increase during NREM sleep and decrease during REM sleep. To address this hypothesis, subjects (n = 19 subjects, see Supplementary Table 1 for subjects' information) participated in Experiment 3 by following the same procedures as those in Experiment 1, except that there was no visual training before or after sleep. The circadian timing of the sleep sessions in Experiments 1 and 3 were equivalent (see Methods).

As shown in Fig. 2a, the mean E/I balance in early visual areas was significantly higher during NREM sleep than at baseline (n = 19 subjects, one sample *t*-test, $t_{18} = 2.91$, $p = 0.009$, Cohen's $d = 0.67$, 95% CI [1.44, 8.87]) in all subjects in Experiment 3. In contrast, during REM sleep, the E/I balance was not significantly different from that at baseline (n = 8 subjects, one sample *t*-test, $t_7 = 1.80$, $p = 0.115$). Next, we directly compared the E/I balance during NREM sleep and REM sleep from Experiments 1 and 3. While the E/I balance during NREM sleep was not significantly different between the experiments (n = 38 subjects, independent-samples *t*-test, $t_{36} = 1.25$, $p = 0.219$), the E/I balance during REM sleep was significantly different between the experiments (n = 18 subjects, independent-samples *t*-test, $t_{16} = 3.50$, $p = 0.003$, Cohen's $d = 1.69$, 95% CI [4.68, 19.04]). These results demonstrate that the E/I balance in early visual areas during NREM sleep increased irrespective of whether presleep learning occurred, while the E/I balance during REM sleep decreased from baseline only after presleep learning.

We further examined the contributions of Glx and GABA concentrations to E/I balance changes during sleep. During NREM sleep (Fig. 2b), while the Glx concentration was not significantly different from baseline in either Experiment 1 or 3 (Exp. 1, n = 19 subjects, one sample *t*-test, $t_{18} = 0.40$, $p = 0.691$; Exp. 3, n = 19 subjects, one sample *t*-test, $t_{18} = 0.40$, $p = 0.695$), the GABA concentration was significantly lower than baseline in both Experiments 1 and 3 (Exp. 1, n = 19 subjects, one sample *t*-tests, $t_{18} = 4.93$, $p < 0.001$, Cohen's $d = 1.13$, 95% CI [−10.18, −4.10]; Exp. 3, n = 19 subjects, one sample *t*-test, $t_{18} = 3.76$, $p = 0.001$, Cohen's $d = 0.86$, 95% CI [−8.11, −2.30], with a Bonferroni adjusted alpha level of 0.0125 (0.05/4)). Thus, the decreases in GABA concentration seem to be a major factor affecting the E/I balance elevation during NREM sleep. However, during REM sleep (Fig. 2c), the

Glx concentration was significantly lower than zero in Experiment 1 (n = 10 subjects, one sample *t*-test, $t_9 = 3.87$, $p = 0.004$, Cohen's $d = 1.23$, 95% CI [−10.57, −2.78], with a Bonferroni adjusted alpha level of 0.01 (0.05/5)), while it was not significantly lower than zero in Experiment 3 (n = 8 subjects, one sample *t*-test, $t_7 = 2.09$, $p = 0.075$), which resulted in a significant difference between Experiments 1 and 3 (independent samples *t*-test, n = 18 subjects, $t_{16} = 4.00$, $p = 0.001$, Cohen's $d = 1.90$, 95% CI [−18.74, −5.77]). However, the GABA concentration during REM sleep was not significantly lower than zero in either Experiment 1 or 3 (Exp. 1, n = 10 subjects, one sample *t*-test, $t_9 = 0.85$, $p = 0.420$; Exp. 3, n = 8 subjects, one sample *t*-test, $t_7 = 0.67$, $p = 0.525$). Thus, the decreases in the Glx concentration seem to be the main determinant of the E/I balance decrease during REM sleep after learning. Because the MRS data quality was comparable between Experiments 1 and 3 (see Quality tests for the MRS data in Methods, Extended Data Fig. 4 and Supplementary Table 11), the differences in the MRS data between the experiments are likely to be caused by the presence of presleep learning.

Experiment 4

Thus far, we have assumed that the preservation of learning of Task-A after training of Task-B is a manifestation of the stabilization of learning of Task-A by REM sleep. However, an alternative interpretation, such as the resource consumption hypothesis, is possible[26]. In this hypothesis[26], the first learning, which consumes all of the plasticity resources during the sleep period, does not allow the second learning to occur; hence, nothing to cause retrograde interference with Task-A should remain after sleep. This suggests that the preservation of learning of Task-A is not caused by its stabilization.

In Experiment 4, we tested whether the resource consumption hypothesis or the stabilization hypothesis that is a part of our complementary processing hypothesis is more plausible (Fig. 3a). The procedures were identical to those of Experiment 2, except that in the pretest and posttest, the performance of both Task-A and Task-B were measured. The resource consumption hypothesis predicts learning of Task-A and no learning of Task-B due to lack of plasticity resources, irrespective of the sleep stages that occur between the tasks. In contrast, the stabilization hypothesis predicts that whether learning occurs depends on whether REM sleep occurs between the tasks. That is, the NREM+REM group should learn both Task-A and Task-B, whereas the NREM-only group should learn neither Task-A nor Task-B due to increased plasticity during NREM sleep and lack of stabilization during REM sleep.

Thus, we divided the subjects into two groups, the NREM+REM group (n = 10 subjects) and the NREM-only group (n = 10 subjects) and tested whether learning of the tasks was dependent on intervening REM sleep. If whether learning occurs depends on whether REM sleep between the two tasks occurs, a significant group difference in the amount of performance improvement in each task should be shown. The performance improvement was defined by [(the performance change between the presleep and postsleep test sessions)/(the performance in the presleep test session)] x 100%, for Task-A and Task-B (See Fig. 3a). A 2-way ANOVA for performance improvement with a Group factor (NREM+REM group vs. NREM-only group) and a Task factor (A vs. B) revealed a significant main effect of

Group (n = 20 subjects, $F_{1,18} = 16.90$, $p < 0.001$, $\eta_p^2 = 0.48$) without an interaction between the factors ($F_{1,18} = 2.23$, $p = 0.153$, see Fig. 3b, c). These results indicate that there was a robust group difference in the performance change for each task, supporting the stabilization hypothesis.

Importantly, the NREM+REM group had significant performance gains for Task-A (the red bar in Fig. 3b; n = 10 subjects, one sample *t*-test, $t_9 = 3.82$, $p = 0.004$, Cohen's $d = 1.21$, 95% CI [9.68, 37.69], with a Bonferroni adjusted alpha level of 0.0125 (0.05/4) here) as well as for Task-B (the blue bar in Fig. 3c; n = 10 subjects, one sample *t*-test, $t_9 = 8.27$, $p < 0.001$, Cohen's $d = 2.62$, 95% CI [33.90, 59.43]). In contrast, the NREM-only group had no performance gains for Task-A or Task-B (gray bars in Fig. 3b, c; Task-A, n = 10 subjects, one sample *t*-test, $t_9 = 0.12$, $p = 0.904$; Task-B, n = 10 subjects, one sample *t*-test, $t_9 = 2.06$, $p = 0.069$). These results support the stabilization hypothesis but are not in accordance with the resource consumption hypothesis.

Discussion

The results of the present study clearly demonstrate two distinctive mechanisms for NREM sleep and REM sleep with complementary patterns of neurochemical and functional processing involved in learning facilitation, as shown in Fig. 4 and Table 1. One mechanism is via plasticity increases shown by the increased E/I balance in early visual areas during NREM sleep. The increased plasticity occurred irrespective of whether presleep learning occurred, that is, in a learning-independent manner. The other mechanism is via stabilization during REM sleep, shown by the decrease in E/I balance to below baseline. In contrast to plasticity increases during NREM sleep, the decreases in E/I balance to below baseline during REM sleep occurred only after presleep learning occurred, that is, in a learning-dependent manner.

The complementary processing model may help resolve the abovementioned three controversies. Instead of assuming only one of the two roles, i.e., off-line performance gain or stabilization, sleep performs both roles in a complementary fashion through different sleep stages: NREM sleep likely plays a role in off-line performance gains, while REM sleep is involved in stabilization. Instead of assuming that the process underlying learning facilitation during sleep is exclusively learning dependent, plasticity increases during NREM sleep occurs in a learning-independent manner, whereas stabilization during REM sleep may occur in a learning-specific manner.

The stabilization that occurs during REM sleep is distinguished from the typical stabilization during wakefulness in several respects. The typical stabilization process during wakefulness takes a few hours[21] and is associated with gradual decreases in the training-induced increased E/I balance back to the baseline E/I balance. On the other hand, the stabilization process during REM sleep is associated with an E/I balance that becomes even lower than baseline in a short time. Interestingly, an E/I balance lower than baseline during wakefulness was observed due to overlearning[21], which is a type of training that is continued even after performance increases reached an asymptote. Overlearning is governed by hyperstabilization[21], which more strongly and rapidly stabilizes a learning state than a

typical stabilization process. Thus, the stabilization process during REM sleep after learning may share some aspects of hyperstabilization during wakefulness.

While MRS measures various types of metabolites outside and inside neurons, the E/I balance changes based on Glx and GABA concentrations measured in our experiments may reflect, if not entirely, synaptic transmission. First, various manipulations that are known to modulate glutamatergic and GABAergic synaptic transmissions have been shown as changes in Glx and GABA concentrations measured by MRS, including drug intake[27], transcranial direct current stimulation[28-30] and memory tasks[31, 32]. Second, both perceptual learning[20, 21, 33] and sleep state[34-36] are known to change glutamatergic and GABAergic activity. Thus, we suggest that the changes in the Glx and GABA concentrations found in the present study may reflect synaptic transmission changes, since these changes in early visual areas were found during conditions that differed only by sleep stages after visual training, with other factors held constant within the same subjects.

The decreases in GABA concentrations in early visual areas during NREM sleep may seem counterintuitive, given that subcortical GABAergic neurons inhibit the ascending arousal system to initiate sleep[36]. One possible explanation is that the activity of these subcortical GABAergic neurons not only initiates sleep but also inhibits local inhibitory interneurons in early visual areas[19, 37]. These disinhibited neurons could enhance plasticity in early visual areas, which leads to off-line performance gains when visual learning occurs before sleep.

The learning-dependent decrease in Glx concentrations during REM sleep may be associated with a reduction in glutamatergic excitatory synapses in neurons activated in early visual areas during preceding NREM sleep. It has been shown that REM sleep selectively prunes and maintains new synapses during learning[38]. Such models of pruning and maintaining synapses in early visual areas are consistent with the role of REM sleep as a stabilizer of presleep learning and enabler of postsleep learning as found in the current study. We assume that the synapses that are involved in visual learning are reactivated during NREM sleep[39] and that the most important synapses will be maintained, and the less important synapses will be pruned during REM sleep. If these pruned synapses are glutamatergic, as excitatory synapses are considered to be the molecular basis of learning[33, 40], pruning of glutamatergic synapses may result in a reduction in the Glx concentration in early visual areas.

One may wonder whether a model of synaptic changes during sleep suggested by a recent review paper[41] is directly related to our findings. Based on the results of animal sleep studies, their model assumes complementary synaptic changes during sleep: local synaptic enhancement during slow wave sleep that occurs specifically for learning and global synaptic downscaling during REM sleep that occurs in a sleep-independent manner[41]. Although their model is highly interesting, it concerns neural processing at a different level/dimension from ours and is not directly linked to our findings. First, the local synaptic enhancement and global downscaling in their model do not have one-to-one correspondence to the off-line performance gain and stabilization, respectively, in our findings. Note that stabilization in our study is defined as the processing of making presleep learning resilient to retrograde interference by new postsleep learning[14]. It is unclear how resilience to interference is formed by their model. Second, in their model, local synaptic enhancement is

assumed to occur in a learning-dependent manner, while global downscaling does not occur specifically for learning. On the other hand, we found that enhanced plasticity linked to off-line performance gain occurs in a learning-independent manner, while hyperstabilization[21], such as decreased plasticity, occurs in a learning-dependent manner. A future study would clarify how the neurochemical and functional changes found in our study are related to their model[41].

A previous study[42] used behavioral measures to reveal that REM sleep eliminates retrograde interference of new learning with old learning. Importantly, both the new and old learning in this study occurred during wakefulness before sleep; hence, presleep interference was eliminated by REM sleep. Although this finding itself is interesting, it is not directly related to the stabilization process of learning during sleep in our study.

It is controversial which spontaneous oscillatory activities in EEG, including those for sigma (13-16 Hz)[43-45] and delta (1-4 Hz)[46] activity during NREM sleep as well as that for theta (5-7 Hz)[47, 48] activity during REM sleep, are involved in learning facilitation during sleep. Therefore, how oscillatory activities relate to the E/I balance during NREM sleep and REM is unclear. Although we cannot draw any strong conclusion from the results (Extended Data Fig. 5) of the present study, the results are not inconsistent with the hypothesis that local sigma activity around early visual areas during NREM sleep may be associated with off-line performance gains, while local theta activity during REM sleep in early visual areas may be associated with stabilization (Extended Data Fig. 5). However, the EEG frequency data were very noisy in the strong magnetic field. Thus, future studies are necessary regarding this matter.

In summary, we examined the roles of NREM sleep and REM sleep in facilitating VPL by measuring neurochemical brain states during sleep in humans. The results of the present study are consistent with the complementary neurochemical and functional processing model, which may lead to resolving the outstanding controversies. The results may also lead to research that will examine how abnormalities in sleep patterns influence the proposed complementary roles of sleep in VPL and other types of learning and memory.

Methods

Subjects

To estimate the number of subjects required to reliably indicate the effect of a daytime nap on off-line performance gains in VPL, we applied the G*Power program[49], with the power set at 0.8 and the required significance level α at 0.05, two-tailed, to our published behavioral data[44] for a t-test. Off-line performance gain in VPL as the dependent variable and t-test as the test family were chosen for the G*Power program, as they were available: There were no studies that tested stabilization of VPL or the relationship between the E/I balance during sleep and performance changes in VPL prior to the present study. The result showed 9 as the sample size. However, to ensure that the results of each experiment were reliable and replicable, we used 19 to 20 subjects for each experiment.

A total of 81 young healthy subjects participated in the present study. See Supplementary Table 1 for the mean age and sex information. All subjects gave written informed consent for their participation in the experiments. This study was approved by the institutional review board at Brown University.

Data exclusion—We removed the data from 5 subjects (see Supplementary Table 1) due to the presence of sleep-onset REM sleep shortly after lights off (n = 4) and ineligibility to participate (n = 1). Thus, the total number of subjects analyzed was 76. See Experimental designs for more details about the classification of subjects into the NREM+REM and NREM-only groups.

Screening process for eligibility—Before consent, the following aspects of the screening process for eligibility took place. First, subjects were eligible when they had normal or corrected-to-normal vision. Second, they were aged between 18-30 years old. Third, based on a self-report, anyone who had a physical or psychiatric disease, was currently using medication, or was suspected to have a sleep disorder was not eligible to participate[50, 51]. Fourth, subjects who had an irregular sleep schedule were not eligible. To test whether subjects had an irregular sleep-wake schedule, they were asked to fill out a sleep-wake habits questionnaire[44, 45], as well as the Munich Chronotype Questionnaire (MCTQ)[52]. The purpose of the sleep-wake habits questionnaire[44, 45] was to identify possible sleep problems. The MCTQ[52] was used to examine personal sleep-wake rhythms on both work and free days. Based on answers to these questionnaires, those who were likely to have problems in their sleep and/or those whose sleep/wake time differed more than 2 hours between weekdays and weekends were regarded as ineligible. Fifth, subjects who had prior experience in visual perceptual learning tasks were regarded as ineligible. Sixth, subjects who played action video games frequently were regarded as ineligible because extensive video game playing affects visual and attention processing[53, 54]. Subjects were asked to fill out a video game questionnaire[44, 45, 54] to report their video game habits. Subjects who had played games for more than 5 hours a week in the past 6 months were regarded as ineligible.

The results of the post-consent questionnaires (see below Common procedures and Supplementary Table 12) indicated that none of our subjects had sleep problems. However, since our screening process was based on a self-report, there was still the possibility that some subjects did not know whether they had sleep disorders. Given this, we excluded subjects if data measured during the experiment showed an irregular sleep structure including irregular sleep-onset REM sleep periods. Thus, at least the sleep structures of our subjects were within a normal range.

Experimental designs

We first describe the procedures common to all experiments and then the procedures specific to each experiment.

Common procedures—The post-consent tests consisted of the Morningness-Eveningness questionnaire (MEQ)[55] and the Pittsburg Sleep Quality Index (PSQI)[56]. The MEQ results indicated that none of the subjects showed extreme degrees of morningness or

eveningness (see Supplementary Table 12). The PSQI results indicated that none of the subjects fell in the category of poor sleepers (see Supplementary Table 12).

Before the experiments started, subjects were instructed to maintain their regular sleep-wake habits, i.e., their daily wake/sleep time and sleep duration, until the study was over. First, the sleep-wake habits of subjects were monitored by a sleep log for approximately a week prior to the experiment. If a subject's sleep/wake time differed by more than 2 hours between weekdays and weekends, a scheduled experiment was postponed. Second, we ensured that subjects had not taken a trip to a different time zone for approximately a week prior to the experiment. Third, all the subjects in Experiment 1 and 18 subjects in Experiment 3 were asked to wear a wrist actigraphy device (GT9X-BT, ActiGraph, Pensacola, FL).

On the day before the experiment, subjects were instructed to refrain from alcohol consumption, unusually excessive physical exercise, and naps. For all experiments, caffeine consumption was not allowed on the day of experiments.

The sleep session started in the early afternoon for all experiments.

In Experiments 1, 2, and 4, there were two training sessions for the texture discrimination task (TDT) (see Texture discrimination task (TDT) for the number of trials), which were separated by a 120-min interval. The background orientations of the presleep TDT (Task-A) and the postsleep TDT (Task-B) were orthogonal to each other (horizontal or vertical). During the 120-min interval, subjects had a 90-min sleep period and a 30-min break. The 30-min break was to ameliorate sleep inertia[57]. During the 30-min break, a questionnaire was administered to obtain subjects' introspection about their sleep, including subjective sleep-onset latency, subjective wake time after sleep onset, comfort of the environment, and occurrence of dreams[51].

In Experiments 1, 2 and 4, subjective (Stanford sleepiness scale; SSS)[58, 59] and behavioral sleepiness (psychomotor vigilance test; PVT[60]) were measured prior to each test session (see Sleepiness measurement for more details). In Experiment 3, there was no test session for the TDT, as there was no learning involved in this experiment (see below). Fourteen out of 19 subjects in Experiment 3 rated their sleepiness using the SSS before electrode preparation, which was approximately 40 min before subjects entered the scanner (see Sleepiness Measurement). This timing roughly matched when the SSS was measured for the subjects in the second test session of Experiment 1.

For all the experiments, there was an additional adaptation sleep session before the main sleep session. When subjects sleep in a sleep laboratory for the first time, the sleep quality is degraded due to the first-night effect (FNE) caused by the new environment[50, 51, 61-63]. The adaptation sleep session was necessary to mitigate the FNE. During the adaptation sleep session, all electrodes for the polysomnography (PSG) measurement were attached to the subjects (see PSG measurement below). Subjects slept in the same fashion as they did in the main sleep session, with (Experiments 1 and 3) or without (Experiments 2 and 4) MRI scans. The adaptation session was conducted approximately one week before the main experimental sleep session so that any effects due to sleeping during the adaptation nap would not carry over to the experimental sleep session.

Two-phase recruitment of subjects—In this study, the recruitment of subjects was performed in two phases in an attempt to equate the number of subjects in the NREM+REM and NREM-only groups. In the first phase, we recruited 10, 30, 9, and 12 subjects for Experiments 1, 2, 3 and 4, respectively. In the second phase, we recruited 30 more subjects. In total, we analyzed 19, 38, 19, and 20 subjects for Experiments 1, 2, 3 and 4, respectively. See Supplementary Table 1 for subject information.

In the first phase, we employed a passive method in which we did not manipulate the sleep structures during the sleep session. That is, we asked subjects to sleep during the sleep session for approximately 90 min. Then, we classified the subjects into the NREM+REM or NREM-only group based on their sleep structure, depending on whether REM sleep appeared or not, during the sleep session. The percentage of the subjects who did not show REM sleep (the NREM-only group) was 20% to 30% in the first phase. Two out of 10 subjects did not show REM sleep (20%) in Experiment 1, 8 out of 30 subjects (27%) did not show it in Experiment 2, 3 out of 9 subjects (30%) did not show it in Experiment 3, and 3 out of 12 subjects (25%) did not show it in Experiment 4.

The second phase was performed to alleviate the imbalance in the number of subjects in the NREM+REM and NREM-only groups in the first phase of all experiments. In the second phase, to specifically increase the number of NREM-only groups, we used a termination method[43]. We terminated the sleep session when the polysomnogram showed a sign of NREM sleep attenuating before the onset of REM sleep. Namely, we terminated the sleep session when sleep stage N1 or N2 reappeared following N3 after NREM sleep lasted more than 40 min and when PSG showed de-synchronized electroencephalogram (EEG) activity and decreased electromyogram (EMG) activity (see PSG measurement below). However, when the attenuation sign appeared before NREM sleep lasted 40 min, we terminated the sleep session so that the REM sleep state did not begin. In each experiment, when the number of subjects for the NREM-only group reached 9, we switched from the termination method to the passive method. See below for the justification for the sleep termination method.

Justification for the sleep termination method—The sleep termination method was justifiable for the following three reasons. First, the results of the previous study that used the sleep termination method[43] replicated the well-documented off-line performance gains in VPL. Second, the length of a sleep session does not affect the frequency drift in MRS data (see Frequency drift in Supplementary Table 11). Third, while it is likely that the sleep termination method makes the sleep session time shorter than the passive method does, the curtailment of the sleep session time does not seem to be associated with any of the behavioral measures or E/I balance in the present study (Supplementary Tables 8-10). We will describe the last point in more detail in below.

In the present study, using data from Experiment 1, we found that sleep time, which is a combination of NREM sleep and REM sleep, was not significantly correlated with off-line performance gain, resilience to retrograde interference, E/I balance during NREM sleep, or E/I balance during REM sleep (see Supplementary Table 8 for detailed statistical results). We also tested whether each NREM sleep duration and REM sleep duration was correlated

with off-line performance gains and resilience to retrograde interference. We found that none of the correlations was significant (Supplementary Table 9).

Next, using data from both Experiments 1 and 3, we computed a correlation coefficient matrix for NREM sleep duration, REM sleep duration, sleep duration (summation of NREM sleep and REM sleep), E/I balance during NREM sleep, and E/I balance during REM sleep. As shown in Supplementary Table 10, none of the coefficients were significantly correlated.

Taken together, we concluded that the termination method was an acceptable procedure in the current study.

Experimental design for Experiment 1—There were four test sessions (Fig. 1b). The first and second test sessions were conducted before and after Task-A training. The third and fourth test sessions were conducted before and after Task-B training.

Shortly after the completion of the first test session, the electrodes for PSG were attached to the subjects (see PSG measurement). Then, the subjects entered the MRI scanner, and PSG started. Before the MRS measurement, an anatomical structure measurement, voxel placement, and shimming were conducted (see MRS acquisition). When the MRS measurement started, the room lights were turned off, and the sleep session began. Approximately 90 min after the PSG recording, the experiment was ended. The MRS measurements were repeated until the experiment ended.

Rough sleep stage scoring was performed online. After noise removal (see PSG measurement), more accurate sleep stage scoring was performed (***Sleep-stage scoring and sleep parameters***). See also MRS acquisition and Co-registration of MRS data and sleep stages for more details.

Experimental design for Experiments 2 and 4—There were two test sessions: pretest and posttest. The pretest session was performed before the first training session on Task-A. After the training session on Task-A, there was a 120-min interval during which the sleep session occurred (see Common procedures above). After the 120-min interval, all subjects performed the second TDT training on Task-B, and the posttest session was conducted.

Eighteen out of 38 subjects were tested only with Task-A in the test sessions, whereas the remaining 20 subjects were tested with both Task-A and Task-B. The data from the latter group of subjects (n = 20) are shown in Experiment 4 (see below).

Experimental design for Experiment 3—After electrodes were attached for the PSG measurement (see PSG measurement), the subjects entered the MRI scanner. After an anatomical structure measurement, voxel placement and shimming for the MRS measurements were conducted (see MRS acquisition). The room lights were turned off, and the sleep session began.

After approximately 90 min of PSG recording, the experiment was ended. After noise removal (see PSG measurement), sleep stage scoring was performed (***Sleep-stage scoring***

and sleep parameters). See also MRS acquisition and Co-registration of MRS data and sleep stages for more details.

Texture discrimination task (TDT)

A TDT, a standard VPL task[24], has been used for studies on the role of sleep in learning[6, 9, 43, 64, 65]. We also used a TDT in Experiments 1, 2 and 4. The TDT was conducted in a dimly lit room. The subject's head and chin were restrained by a chin rest to precisely maintain a viewing distance of 57 cm. Stimuli were generated by MATLAB software with the psychtoolbox[66, 67].

The TDT stimulus (Fig. 1a) consisted of a target array of three-line segments aligned vertically or horizontally with the background line segments. The TDT consisted of two tasks: the orientation task and the letter task. The orientation task was the main task, whereas the letter task was designed to control subjects' fixation[24]. Each trial (Fig. 1a) began with the presentation of a fixation point at the center of the screen (1000 ms). Then, a target display was briefly presented (17 ms), followed by a blank screen of varying duration and then a mask stimulus (100 ms), which was composed of randomly rotated v-shaped patterns. Subjects were instructed to fix their eyes on the center of the display throughout the stimulus presentation.

The size of the target display was 19° and contained a 19 x 19 array of background lines. Each background line was jittered by 0.2°. The orientation of the background lines was either horizontal or vertical. Each target display had 2 components: a letter (either 'L' or 'T') presented at the center of the display and 3 diagonal lines presented at a peripheral location within a trained visual field quadrant at 5–9° of eccentricity. The target array of three lines was aligned either horizontally or vertically on the background. The location of the target array was either in the left or right upper visual field quadrants, randomly assigned to each subject, and kept consistent throughout the experiment. After the mask display, subjects used a keyboard to report whether the central letter was 'L' or 'T' (the letter task) and whether the target array was aligned 'horizontally' or 'vertically' (the orientation task). After the subject's responses for the letter and orientation tasks, a feedback sound was delivered to indicate whether the letter task was correct (1000 Hz pure beep) or incorrect (400 Hz pure beep). No feedback was given for responses on the orientation task.

The time interval between the target onset and mask is referred to as the stimulus-to-mask onset asynchrony (SOA). The SOA was modulated across trials to control the task difficulty. The task difficulty increased as the SOA decreased. There were 5 or 6 SOAs ranging from 316-33 ms in an experiment. In a test session, there were 15 or 20 trials for each SOA. Thus, there were a total of 75 or 120 trials in a test session. The presentation order of the SOAs was pseudorandomized to reduce the amount of learning and fatigue during the session[68]. One test session took approximately 6-10 min. In a training session, 60 trials were blocked by SOA. The order of the SOA presentation was in a descending manner. Training for the presleep TDT (Task-A) was conducted before sleep with one or two blocks per SOA used, for a duration of approximately 20-40 min. Training for the postsleep TDT (Task-B) was conducted after sleep with one block per SOA and took approximately 20 min. The orthogonal orientations (horizontal or vertical) of the background lines were used for Task-A

and Task-B. One of horizontal or vertical background orientations was randomly assigned to Task-A, whereas the other orientation was assigned to Task-B. Within a test or training session, there was a 10-s break every 15 or 20 trials. Between sessions, there was a 2-min break.

The performance of the TDT was defined as the threshold SOA at which a subject marks 80% correct responses. The threshold SOAs were obtained for each test session for each subject in the following way. The percentage of correct responses for the orientation task was calculated for each SOA in a test session. A cumulative Gaussian function was fitted to obtain a psychometric curve to determine the threshold SOA that corresponded to 80% correct performance using the psignifit toolbox (ver. 2.5.6) in MATLAB (http://bootstrap-software.org/psignifit;[69]). All trials in which the letter task was incorrect were removed from the calculation of the threshold SOA.

To explain how to perform the TDT to a naive subject in a standardized manner, we had an introductory session before the first test session in the experiments. During the introductory session, 3 long SOAs (800, 600, 400 ms) were used. The introductory session started with the longest SOA (800 ms) and was followed by 600-ms and 400-ms SOAs in a blocked fashion, where a set of 15-20 trials was conducted for each SOA (a total of 45 or 60 trials). The introductory session was repeated until the subject performed the orientation task with no or only one incorrect response for the 400-ms SOA trials.

Only in Experiment 1 was there a reminder session before the third test session to remind the subjects how to conduct the task after the 2-hr time interval. The reminder session had the same 3 sets of SOAs as the introductory session. We did not conduct a reminder session in Experiment 2 or 4 due to temporal proximity between the posttest and the training session for Task-B.

TDT performance changes were computed based on relative changes in the threshold SOAs (ms) among test sessions. The performance change (%) at a test session was calculated as [(the threshold SOA change between a previous and the current test session)/(the threshold SOA at a previous session)] x 100%.

Sleepiness measurement

The SSS rating[58, 59] ranged from 1 (feeling active, vital, alert, or wide awake) to 7 (no longer fighting sleep, sleep onset soon; having dream-like thoughts). Subjects were asked to choose the scale rating that described their state of sleepiness.

The PVT was implemented with open-source Psychology Experiment Building Language (PEBL) software[60]. In a trial of the task, after a fixation screen, a target screen was presented in which a red circle appeared on the center of the screen. The subjects were required to press the spacebar on a keyboard as quickly as possible upon detection of the circle. The time interval between the fixation screen and the screen with the red circle varied between 1000–10000 ms. The PVT lasted for approximately 2 min. The reaction time (RT, s) was log-transformed to reduce the skew of the data[70]. The average RT, the number of lapses, the 10% fastest RTs, the 10% slowest RTs, and the lapse threshold were obtained as

measurements of behavioral sleepiness (see Supplementary Table 3-6 for sleepiness data), according to a previous study[71].

Sleepiness data for each experiment, including the SSS score and 5 measures obtained from the PVT (mean reaction time (RT), the number of lapses, 10% fastest RTs, 10% slowest RTs and the lapse thresholds), were collected using procedures from a previous study[71]. The lapses, defined as RTs greater than 500 ms, were measured for each subject and for each test session. The lapse thresholds at 500 ms were obtained by plotting a cumulative distribution function of reaction times across subjects for each group.

We tested whether each sleepiness measure was different between the NREM+REM group and NREM-only group in each experiment. The Shapiro-Wilk test was used to test the normality of the data. For the mean RTs in Experiments 1 and 3, because the data were regarded as being normally distributed, we performed a mixed-design ANOVA with Group as a between-subject factor and Test session as a within-subject factor. Mauchly's test of sphericity showed no violation of sphericity. For the other measures, including the SSS scores, mean RTs in Experiment 2, and 10% fastest and slowest RTs, the Mann-Whitney U test was used for each test session because of the violation of normality. There were no significant differences in SSS scores or any PVT measures between the groups in any experiment. See Supplementary Tables 3 - 6 for the statistical results.

PSG measurement

Experiments 1 and 3: Concurrent PSG and MRS—It took approximately 30 min to attach electrodes for PSG[1, 72]. PSG was obtained simultaneously with structural MRI and MRS. PSG consisted of an electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), and electrocardiogram (ECG). For all subjects except for 2 subjects in Experiment 1, EEGs were recorded at 23 scalp sites using an MRI compatible 32-channel EEG cap (32Ch BrainCap MR with Multitrodes, Brain Products GmbH, Gilching, Germany). Due to temporary unavailability of the 32-channel EEG cap, for the 2 subjects in Experiment 1, EEGs were recorded at 18 scalp sites using an MRI compatible 25-channel EEG cap (25Ch BrainCap MR with Multitrodes, Brain Products GmbH, Gilching, Germany). EOGs were recorded with 2 electrodes bipolarly placed at the outer canthi of both eyes (horizontal EOG) and with 4 electrodes placed above and below the left and right eyes (vertical EOG) when a 32-channel EEG cap was used. When a 25-channel EEG cap was used, EOGs were recorded with 2 electrodes bipolarly placed at the outer canthi of each eye and with 2 electrodes placed above and below the left eye. EMGs were recorded bipolarly from the mentum. ECGs were recorded from the lower shoulder blade. EEG and ECG were referenced to Fz. The ground electrode was placed at AFp4. Electrode impedances were kept approximately at or below 5 kΩ for EEG and ECG and at approximately 10 kΩ for EOG and EMG. All data were recorded by an MRI-compatible amplifier (BrainAmp MR, Brain Products GmbH, Gilching, Germany) and a recording software (BrainVision Recorder, Brain Products GmbH, Gilching, Germany) at a sampling rate of 5000 Hz.

PSG recorded simultaneously with MRI contains two types of noise: scanner noises and ballistocardiogram artifacts. First, scanner noise was removed after the experiment using

Brain Vision Analyzer 2 (Brain Products GmbH, Gilching, Germany). Next, to remove ballistocardiogram artifacts, the data were low-pass filtered at 100 Hz and downsampled to 250 Hz. Then, the FMRIB plug-in for EEGLAB (The University of Oxford) was used. EEG data were re-referenced to TP9 and TP10 using EEGLAB.

See Co-registration of MRS data and sleep stages below for details about how we co-registered the MRS data and sleep stages.

Experiments 2 and 4: PSG only—It took approximately 45 min to attach electrodes for the PSG[1, 72]. PSG consisted of EEG, EOG, and EMG. EEG was recorded at 24 scalp sites according to the 10-10 system for electrode positioning[73] using a cap (actiCAP, Brain Products GmbH, Gilching, Germany) and active electrodes (Brain Products GmbH, Gilching, Germany) with a standard amplifier (BrainAmp Standard, Brain Products GmbH, Gilching, Germany). The reference, which was Fz online, was measured with active electrodes and re-referenced to the average of the left (TP9) and right (TP10) mastoids after the recording for analysis. The sampling frequency was 500 Hz. The impedance was kept below 20 kΩ. The active electrodes included a new type of integrated impedance converter, which allowed them to transmit the EEG signal with significantly lower levels of noise than traditional passive electrode systems. The data quality with active electrodes was as high as 5 kΩ using passive electrodes[51]. The passive electrodes (BrainAmp ExG, Brain Products GmbH, Gilching, Germany) were used for EOG and EMG. Horizontal EOG was recorded using two electrodes placed at the outer canthi of both eyes. Vertical EOG was measured using 2 electrodes 3 cm above and below the left eye. EMG was recorded from the mentum (chin). The impedance was maintained at approximately 10 kΩ for the passive electrodes. Brain Vision Recorder software (Brain Products, LLC) was used for recording. The data were filtered between 0.1 and 100 Hz. PSG was recorded in a soundproof and shielded room.

Sleep-stage scoring and sleep parameters

In Experiments 1 and 3, when the sleep session took place in the MRI scanner, PSG data were segmented for the purpose of co-registration with MRS data. See Co-registration of MRS data and sleep stages below for more details. In Experiments 2 and 4, no further segmentation was performed for the PSG data.

Sleep stages were scored for every 30-s epoch, following the standard criteria[1, 72], into stage wakefulness (stage W), NREM stage 1 sleep (stage N1), NREM stage 2 sleep (stage N2), NREM stage 3 sleep (stage N3), and stage REM sleep (REM sleep). Sleep staging was performed for each PSG segment in Experiments 1 and 3.

Standard sleep parameters were obtained (Supplementary Table 2) to indicate a general sleep structure for each experiment. Sleep parameters included the sleep-onset latency (SOL, the latency to the first appearance of stage N2 from lights off), the percentage of each sleep stage, the wake time after sleep onset (WASO), sleep efficiency (SE, the total percentage spent in sleep), and the time in bed (TIB, the time interval between lights off and lights-on)[72]. Only the data from the first sleep cycle were included in the analysis. Standard sleep parameters were obtained based on the PSG segments in Experiments 1 and 3.

MRS acquisition

Subjects were scanned using a 3T Siemens Prisma scanner (Siemens) with a 64-channel head coil. It was important for subjects to sleep without discomfort and head motion during the MRI measurements. Cushions and gauze were used to stabilize subjects' heads to reduce discomfort, and we ensured that there would be no space left between subjects' heads and the head coil to reduce head motion. A thin back cushion and knee cushion were used upon the subjects' request. Several blankets were used to keep the subjects warm and to initiate sleep during the scan.

For all MRS experiments, an anatomical (T1-weighted) reference image dataset was acquired using a magnetization prepared rapid gradient echo (MPRAGE) sequence (256 slices, voxel size=1x1x1 mm, 0 mm slice gap) to localize the voxel of interest (VOI) for MRS. Based on the measured anatomical structure, the VOI was manually placed on the most posterior part of the occipital lobe, covering the calcarine sulci that corresponds to early visual areas bilaterally[21]. We carefully placed the VOI to include the least amount of white matter, as lipids in the white matter cause noise in the spectra. Shimming for MRS was then performed using a vendor-provided automated tool (defined by the full width at half maximum of the water peak; see Quality tests for the MRS data). The procedure for the structural measurement, VOI placement, and shimming took approximately 20-30 min.

The MEGA-PRESS sequence[74-76] (see below) was used to measure concentrations of both GABA and Glx simultaneously from the VOI so that GABA and Glx were acquired from the same scan during the same sleep stages, according to the procedure used in previous studies[77-82]. We used a shorter TR and a larger VOI size than a previous study[21] to increase the signal-to-noise ratios.

For all the subjects in Experiment 1 and 14 subjects in Experiment 3, we first conducted a quick test scan of the MEGA-PRESS sequence (TR = 1.25 s, TE = 68 ms, number of average = 32, for edit-on and edit-off, see below) with double-banded pulses to simultaneously suppress the water signal[83] and edit the γ-CH2 resonance of GABA at 3.0 ppm[21, 74, 76, 84] to check whether the spectrum was acceptable. Second, one unsuppressed water spectrum was acquired (TR = 1.25 s, TE = 68 ms, number of average = 16) as a standard water concentration reference for single-voxel proton MRS[85-87]. Third, the 5-s dummy scans for the steady state of longitudinal magnetization and 10-min water-suppressed MEGA-PRESS sequence (TR = 1.25 s, TE = 68 ms, number of average = 240 for each edit-on and edit-off, VOI = 2.2×2.2×2.2 cm^3) were run repeatedly until the sleep session was over.

For the remaining 5 subjects in Experiment 3, each of the acquisition times for the MEGA-PRESS scans with water suppression was 3.3 min (TR = 1.5 s, TE = 68 ms, number of average = 64, for each edit-on and edit-off, VOI = 2x2x2 cm^3, scan time 198 s including 6-s dummy scans for the steady state of longitudinal magnetization). None of the quick test scans or water-unsuppressed scans were performed. The spectra were obtained from the 3.3-min segment.

For all the data in Experiments 1 and 3, the final spectra were obtained by subtracting the signals from alternate scans with the selective double-banded pulse applied at 4.7 and 7.5 ppm (edit-off) and the selective double-banded pulse applied at 1.9 and 4.7 ppm (edit-on). The differential spectrum showed only the outer lines of the GABA triplet at 3.0 ppm, as the creatine singlet would cancel it out. The bandwidth of the frequency-selective pulses permitted assessment of N-acetylaspartic acid (NAA) and glutamate resonance at 3.7 ppm.

Each of the segmented spectroscopic data(see Co-registration of MRS data and sleep stages below for more details about how the MRS data were segmented to co-register with sleep stage) was processed using LCModel[88, 89] for metabolite quantification of neurotransmitters, including Glx, GABA, and NAA. Note that Glx is a combined signal from glutamate with glutamine. The contribution of glutamine to Glx is less than 15%[32]. The LCModel assumes that the obtained spectrum can be fitted in the frequency domain using a linear combination of basis functions. The amounts of GABA and Glx were normalized by (i.e., divided by) the amount of NAA and referred to as the concentrations of GABA and Glx, respectively[30, 78, 82, 90].

Co-registration of MRS data and sleep stages

We segmented both PSG and MRS data so that MRS data and sleep staging would temporally co-register in Experiments 1 and 3. The co-registration procedures were different for those for whom MRS was performed with a 10-min sequence (all the subjects in Experiment 1 and 14 subjects in Experiment 3) and those for whom MRS was performed with a 3.3-min sequence (the remaining 5 subjects in Experiment 3).

1) For all the subjects in Experiment 1 and 14 subjects in Experiment 3

1-1. MRS segments: MRS was performed every 10 min. The MRS data for each 10 min were termed an “MRS segment”. For example, “Segments” #1 to #9, in Extended Data Fig. 6a, each corresponds to one 10-min MRS segment. The measurement of each MRS segment was manually started by an experimenter (see MRS acquisition above).

1-2. PSG measurement: The PSG measurement was conducted for approximately 90 min without any breaks, as mentioned above (see Experimental design for Experiment 1, Experimental design for Experiment 3, and PSG measurements). Sleep scoring was performed for every 30-s epoch in the standardized way[1, 72] (see Sleep-stage scoring and sleep parameters above and Extended Data Fig. 6a).

1-3. MRS segment and PSG data alignment: The scanner noise caused by MRS was easily detected on the PSG record. This allowed us to temporally align the start of each MRS measurement on the PSG record. Blue dots in Extended Data Fig. 6a illustrates sleep stages scored every 30 s. Note that there was a small time interval (up to 10 s) between the end of one MRS measurement and the start of the next measurement. The PSG records between the MRS measurements were omitted so that the lengths of the MRS segments and PSG segments were matched. The MRS segments and sleep stages in Extended Data Fig. 6a do not show the omitted PSG records.

1-4. Combining N1, N2, and N3 stages as NREM sleep: Since sleep stages N1, N2, and N3 are subcategorized stages of NREM sleep, we combined sleep stages N1, N2 and N3 together and labeled them NREM sleep in the present study. Thus, Extended Data Fig. 6b shows that only 3 PSG states (NREM sleep, REM sleep and Wakefulness) were registered.

1-5. Further segmentation: As shown in 'New' in Extended Data Fig. 6c, some of the 10-min MRS-PSG segments were further split into five 2-min segments. We determined whether we should perform further segmentation based on whether each 10-min PSG segment had a single PSG stage or multiple PSG stages (see Extended Data Fig. 6b).

1-5-1. For dominance of a single sleep stage in a PSG segment.: When a 10-min PSG segment was dominated by a Wakefulness, NREM sleep, or REM sleep state (i.e., more than 80% of the segment, as in the original segments #2-#7 in Extended Data Fig. 6c), its corresponding MRS segment was registered as that PSG state (Extended Data Fig. 6c). No further segmentation was conducted for the MRS-PSG segments.

1-5-2. For nondominance of a single sleep stage in a PSG segment.: When the Wakefulness, NREM sleep, or REM sleep states in a 10-min PSG segment did not exceeded 80% of the segment, as in the original segment #1 or #8 in Extended Data Fig. 6c, the segment was further split into five 2-min MRS segments using the raw data (twix files), as in the new segments #12-#16 in Extended Data Fig. 6c. The 10-min PSG segment was also split into five corresponding 2-min PSG segments. Thus, each 2-min MRS segment was coupled with its corresponding 2-min PSG segment.

1-6. Co-registration of MRS and PSG for 2-min segments: A 2-min PSG segment consisted of 4 sleep stage epochs, as a sleep stage was scored every 30 sec as described in the above PSG measurement section. Each 2-min MRS segment was registered to be in the wakefulness, NREM sleep or REM sleep state by taking the statistical mode (the most frequent) of the 4 sleep stage epochs. If there were the same numbers of different sleep stages in a 2-min PSG segment, the priority registration occurred in the following order: Wakefulness > NREM sleep > REM sleep. For instance, if one 2-min PSG segment contained 2 Wakefulness epochs and 2 NREM sleep epochs, the corresponding MRS segment was registered to be during Wakefulness. The priority ordering increases the stringency of the registration of NREM sleep and REM sleep. This ordering was conducted for the following two purposes. The first purpose was to remove any MRS segments containing a Wakefulness stage from MRS segments registered to be during sleep stages. This reduced possible contamination of the MRS data for a sleep state with that for a Wakefulness state. The second purpose was to obtain MRS data for REM sleep with a reduced possibility of contamination by an NREM sleep state. An MRS segment was removed from further analyses if either the number of PSG epochs containing noises or the number of epochs showing arousal[72] (that is, a short burst of alpha or theta activity that is not long enough to be scored as Wakefulness) was equal to or more than half of all the epochs in a PSG segment, because that MRS segment was too unreliable to be classified into a particular sleep stage. See Supplementary Table 13 for the average MRS segments.

2) The remaining 5 subjects in Experiment 3—Since no twix files were stored, there were no further changes in MRS segmentation for these data. One 3.3-min MRS segment contained six corresponding 30-sec sleep stage epochs. A Wakefulness, NREM sleep or REM sleep state was registered to be in one 3.3-min MRS segment by taking the mode (highest frequency) of the sleep stages that occurred during the segment. If the same numbers of different sleep states were included in a 3.3-min MRS segment, (1) wakefulness was prioritized over NREM sleep and REM sleep, and (2) NREM sleep was prioritized over REM sleep, as mentioned above.

3) Calculation of the mean E/I balance for NREM and REM sleep—The calculation of the mean E/I balance for NREM and REM sleep was the same for all subjects. After the co-registration between the MRS segments and PSG sleep stages, the E/I balance for each MRS segment ($E/I_{segment}$) was calculated by dividing the Glx concentration by the GABA concentration. Then, the mean E/I balance during wakefulness ($E/I_{baseline}$) was calculated as the baseline. Note that there was no significant difference between the E/I balance during wakefulness before and after sleep onset ($U = 44$, $p = 0.678$, $N_{wake\text{-}before} = 10$, $N_{wake\text{-}after} = 10$, using data in Experiment 1). Thus, the mean E/I balance during wakefulness was taken from the E/I balance both before sleep onset and after sleep onset. Next, the E/I balance for each MRS segment was normalized by $[(E/I_{segment} - E/I_{baseline})/(E/I_{baseline})] \times 100\%$. Finally, the mean of the normalized E/I balance was calculated for each NREM sleep and REM sleep. See Extended Data Fig. 6d and e.

Quality tests for the MRS data

We used a relatively short MRS segment to extract a spectrum. We conducted 4 types of quality checks, including analyses for shim values, NAA linewidth, frequency drift for the spectra, and Cramer-Rao lower bounds, as shown in Supplementary Table 11. We compared the E/I balance between Experiments 1 and 3 to test whether presleep learning affected the E/I balance during sleep. To check whether this was a fair comparison, we tested whether the MRS data quality shown in these measures were comparable between Experiments 1 and 3. The results (see below) indicate that there was no systematic bias in the MRS data quality between Experiments 1 and 3. Thus, the MRS data in Experiments 1 and 3 were sufficiently reliable for a fair comparison.

Shim values.—The shim values represent the homogeneity of the magnetic field, which was obtained once for each MRS run for each subject. A shim value of less than 30 Hz is regarded as desirable[91]. In the present study, the mean shim value was much lower than 30 Hz and comparable to a previously reported value[79].

NAA linewidth.—The linewidth for NAA, which determines the resolution available to discern spectral features (therefore, the lower the better), was noted for each MRS run for each subject. We confirmed that the NAA linewidth was below 10 Hz for all subjects (see Extended Data Fig. 2 for example spectra and Extended Data Fig. 4 for the NAA linewidth). The mean NAA linewidth for each experiment was comparable to values reported in previous studies[21, 79].

Frequency drift.—We measured the frequency drift, in which larger values suggest more head motion. The mean frequency drift values were in an acceptable range, similar to those from previous studies[21, 79]. There were no frequency drift data from the 5 subjects in Experiment 3. Thus, the frequency drift in Experiment 3 was calculated from 14 subjects.

Cramer-Rao lower bounds.—The Cramer-Rao lower bounds (or %SD) were used as a measure of fitting error (therefore, the lower the better). A commonly accepted Cramer-Rao lower bound criterion of 20% was chosen to reject low-quality signals[92]. The mean %SD values for Glx and GABA were similar to values reported in previous studies[21].

Comparison of the MRS quality in Experiments 1 and 3.—To test whether MRS data quality was significantly different between Experiments 1 and 3, the Shapiro-Wilk test was first used to test the normality of data for each of the shim values, NAA linewidths, frequency drifts, and %SDs. For the shim values, because the Shapiro-Wilk test indicated that the data distribution was normal, we conducted a two-sided independent-samples *t*-test between Experiments 1 and 3. For other values, a two-sided Mann-Whitney U test was conducted because of the violation of normality. None of the results showed a significant difference between Experiments 1 and 3: the shim value ($t(36) = 0.70$, $p = 0.490$), the NAA linewidth averaged across runs ($U = 169$, $p = 0.748$), frequency drift of the first run ($U = 119$, $p = 0.623$), frequency drift of the last run ($U = 85$, $p = 0.084$), %SD for Glx (NREM sleep: U=124, $p = 0.102$; REM sleep: $U = 20$, $p = 0.080$) or GABA (NREM sleep: $U = 166$, $p = 0.683$; REM sleep: $U = 31.5$, $p = 0.472$).

Comparison of the MRS quality between NREM sleep and REM sleep.—To test whether the MRS data quality was significantly different during NREM sleep and REM sleep, a two-sided Wilcoxson signed-rank test was conducted for frequency drifts, and a two-sided Mann-Whitney U test was conducted for %SDs because of the violation of normality shown by the Shapiro-Wilk test. We first tested whether the frequency drifts were significantly different between the first run vs. the last run of the MRS scans for Experiments 1 and 3. Because REM sleep would occur towards the end of the scans, while NREM sleep would occur in the beginning of the scans, a significant difference in the frequency drifts between the first and last runs, if any, would suggest that there were MRS quality differences between the sleep stages. However, there was no significant difference in the frequency drifts between the first and the last runs for Experiment 1 ($Z = 0.56$, $p = 0.573$) or Experiment 3 ($Z = 0.41$, $p = 0.683$). We next tested whether the %SD was significantly different during NREM and REM sleep for both Glx and GABA concentrations in Experiments 1 and 3. None of the % SDs for Glx concentration (Exp. 1: $U = 52.5$, $p = 0.053$; Exp. 3: $U = 46.5$, $p = 0.123$) or GABA concentration (Exp. 1: $U = 71.5$, $p = 0.289$; Exp. 3: $U = 37$, $p = 0.041$, with a Bonferroni adjusted alpha level of 0.025 (0.05/2)) were significantly different between the sleep stages. These results indicate that overall, the MRS data quality showed no significant difference during NREM sleep and REM sleep. Note that the %SD for GABA was significantly different during NREM sleep and REM sleep in Experiment 3 (without a correction for multiple comparisons). However, because the %SDs for GABA concentration during each sleep stage (7.9% for NREM sleep, 9.0% for REM sleep) in Experiment 3 were well below the recommended threshold (20%)[92], these values

were acceptable. Importantly, there were no other results that showed a significant difference during NREM sleep and REM sleep in Experiment 3. Thus, this difference in %SD for GABA during NREM sleep and REM sleep in Experiment 3 is unlikely to explain the significant difference in the Glx concentrations and E/I balance during REM sleep between Experiments 1 and 3.

Statistical analyses

An α level (Type I error rate) of 0.05 was set for all statistical analyses. The Shapiro-Wilk test was conducted for all the data to test whether the data were normally distributed. If data were not normally distributed, nonparametric tests including the Mann-Whitney U test and Kruskal-Wallis test were used. When normality was not rejected by the Shapiro-Wilk test, Levene's test was conducted to test for the homogeneity of variance before conducting parametric tests including one sample *t*-tests, independent-samples *t*-tests, Pearson's correlation analysis, and ANOVA. When a correction for multiple comparisons was needed, the Bonferroni correction was applied and the adjusted alpha level was shown, but all *p* values are reported without any adjustments.

All tests conducted in this study were two-tailed. When a test indicated a statistical significance, the effect size was shown using Cohen's d^{93} for *t*-tests, and η_p^2 for a main effect of ANOVA. We also included a 95% CI for a significant *t*-test and Pearson's correlation analysis.

Statistical tests were run by SPSS (ver. 24, IBM Corp.) and MATLAB (R2016b and R2019b, MathWorks, Inc.).

Randomization—In TDT training (see Texture discrimination task (TDT) above), proper randomization for visual stimuli was performed. Pairs of background orientations for Task-A and -B were randomized across subjects. Also, the the visual quadrant where the target appeared in the task was randomly assigned to upper left or upper right for each subject.

Blinding—Data collection and analysis were not performed blind to the conditions of the experiments.

Statistics and Reproducibility—Behavioral results of Exp.1 were replicated in Exp. 2 and 4.

Reporting Summary

Further information on research design is available in the Life Science Reporting Summary linked to this article.

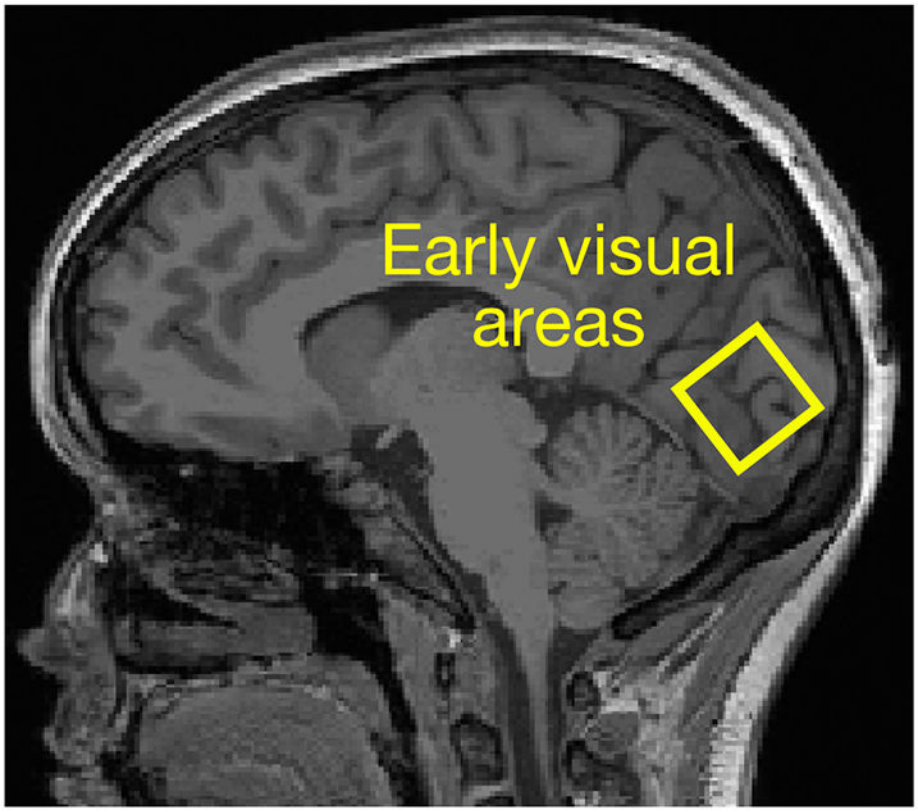
Data availability

The datasets generated and/or analyzed in the current study are included in this published article as Source Data and Supplementary Table 14.

Code availability

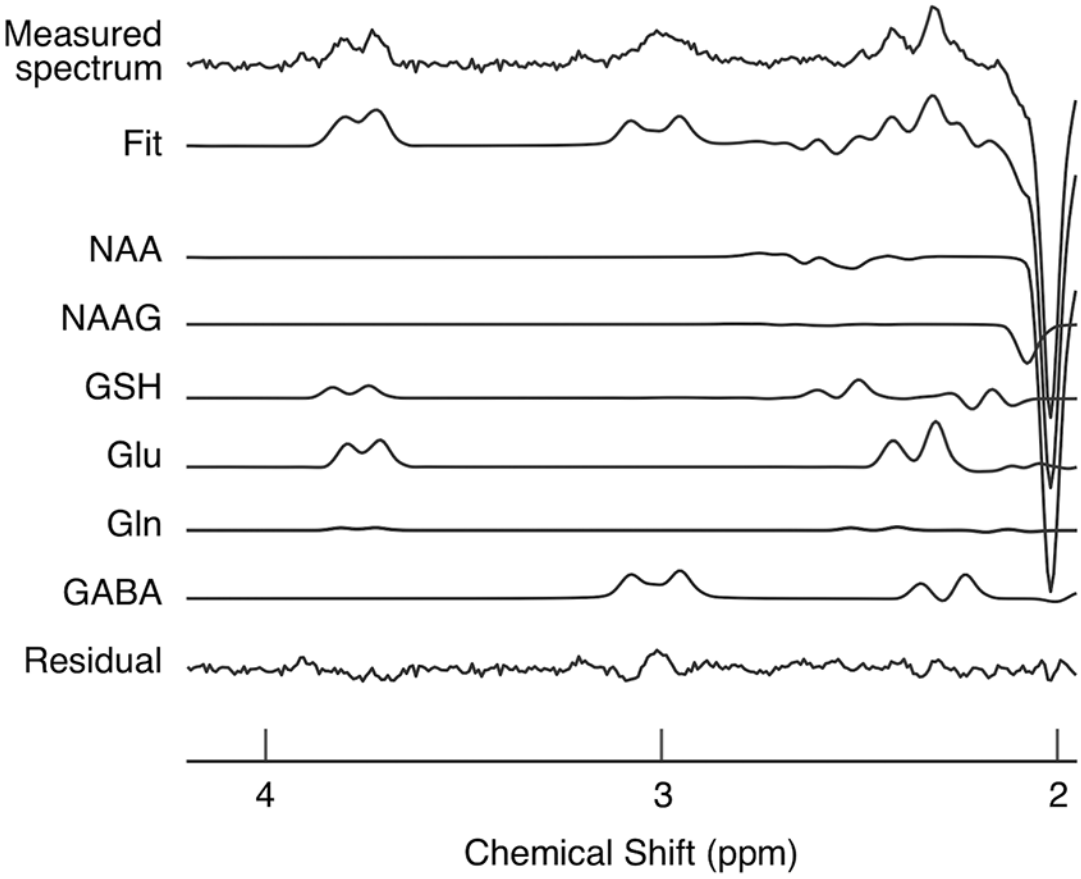
The computer code that was used to generate results central to the conclusions of this study is available from the corresponding author upon request. See Supplementary Software for the software list.

Extended Data

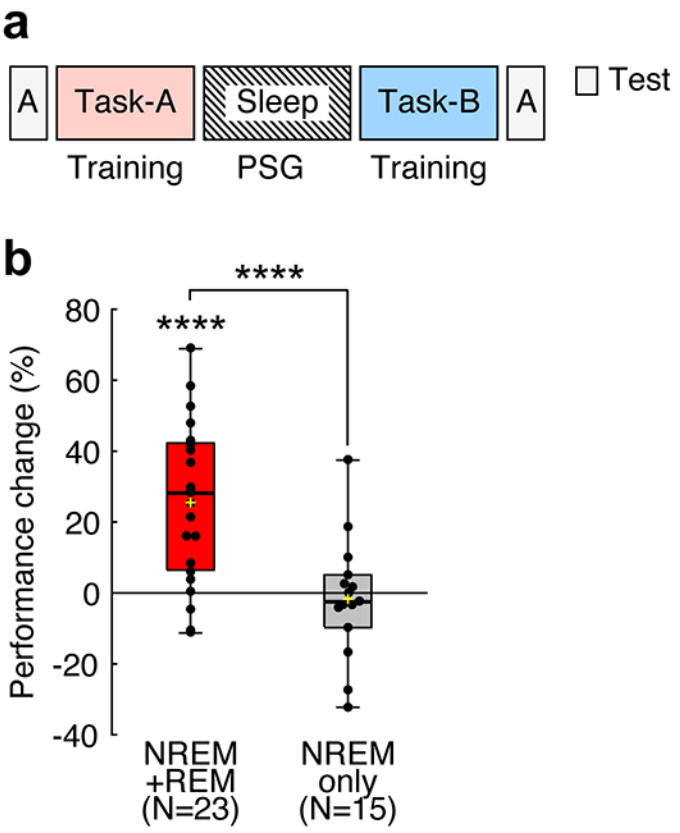


Extended Data Fig. 1. Example structural MRS image indicating the voxel located in early visual areas.

Based on the measured anatomical structure, the voxel of interest was manually placed on the most posterior part of the occipital lobe, covering the calcarine sulci that corresponds to early visual areas bilaterally[21].

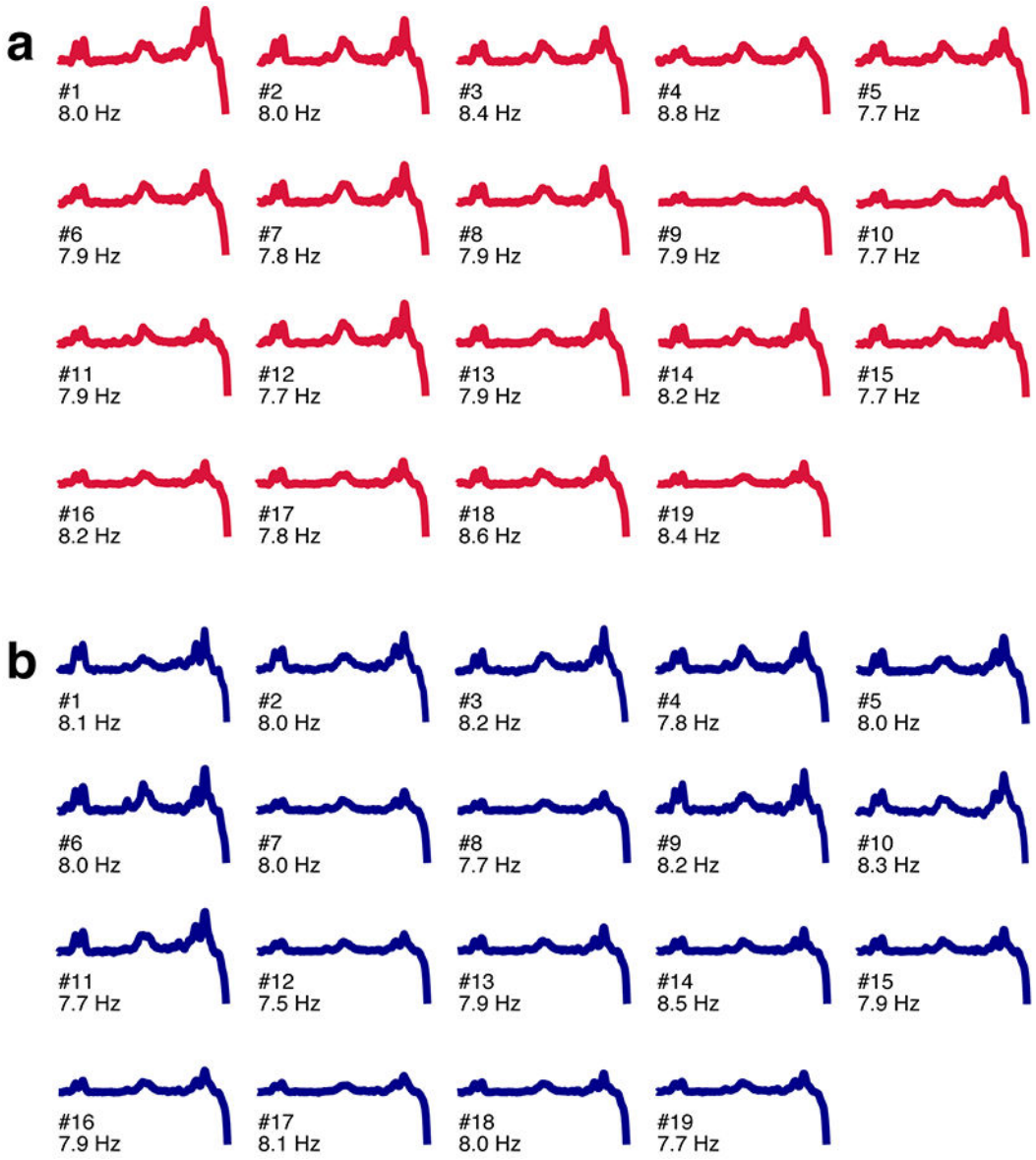


Extended Data Fig. 2. Example spectra from the voxel located in early visual areas.
The measured spectrum is shown in the top row. “Fit” in the second row represents the spectrum fitted with the LCModel (see MRS acquisition in Methods for details). “Residual” in the bottom row represents the residual remaining after the fitting. The remaining rows indicate individual fits for all metabolites that can be detected by a given acquisition. Macromolecular and lipid signals were used for the baseline correction. NAA, NAAG, GSH, Glu, Gln, and GABA represent N-acetylaspartate, N-acetylaspartylglutamate, glutathione, glutamate, glutamine, and gamma-aminobutyric acid, respectively. Glx is obtained by adding glutamine and glutamate in the LCModel. The same procedure was repeated 272 times independently with similar results.



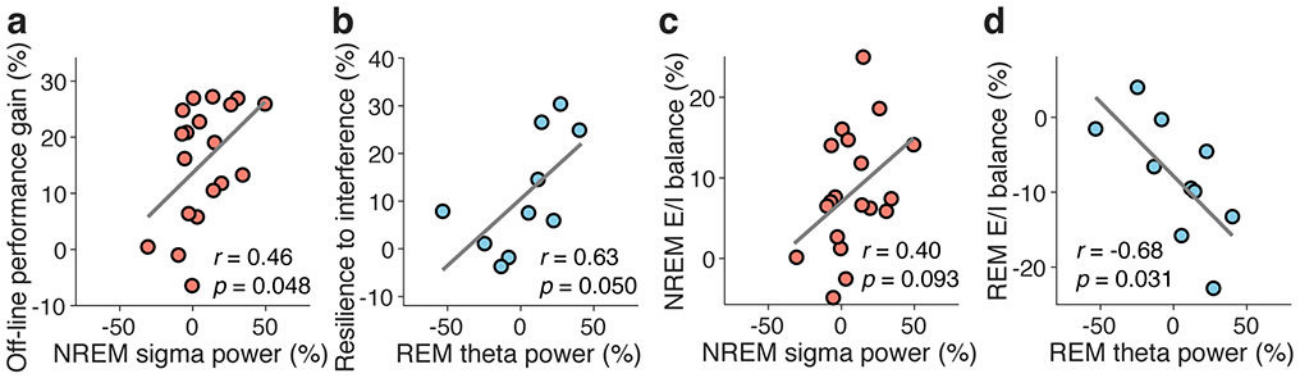
Extended Data Fig. 3. The experimental design and results of Experiment 2.

The experimental design and results of Experiment 2. **a**, Design. The first texture discrimination task (TDT) training (Task-A) was conducted with background A, and the second training (Task-B) was conducted with background B. Test sessions were conducted before the first training (pretest) and after the second training (posttest) to measure performance gains on Task-A. **b**, Boxplots for the performance change (%) for Task-A from the pretest session to the posttest session in the NREM+REM group (red box, n = 23 subjects) and the NREM-only group (gray box, n = 15 subjects). Significant off-line performance gains were observed among subjects who had both NREM sleep and REM sleep (**** two-sided one sample *t*-test against 0, $t_{22} = 5.48$, $p < 0.001$, $d = 1.14$, 95% CI [15.86, 35.17]), whereas the subjects who had only NREM sleep did not show any significant off-line performance gains (two-sided one sample *t*-test against 0, $t_{14} = 0.39$, $p = 0.699$). Furthermore, off-line performance gains were significantly different between the groups (**** two-sided independent-samples *t*-test, $t_{36} = 4.02$, $p < 0.001$, $d = 1.33$, 95% CI [13.49, 41.01]). For each boxplot, the bottom and top of the box correspond to the 25th and 75th percentiles (the lower and upper quartiles), respectively. The inner thick horizontal line represents the median, and the plus mark represents the mean. The whiskers show the maximum and minimum of the data. Individual data (dots) are overlaid. Grubbs' test showed no outliers (Alpha=.05, two-sided).



Extended Data Fig. 4. Mean raw spectra in Experiment 1 (a, n=19 subjects, red plots) and in Experiment 3 (b, n=19 subjects, blue plots).

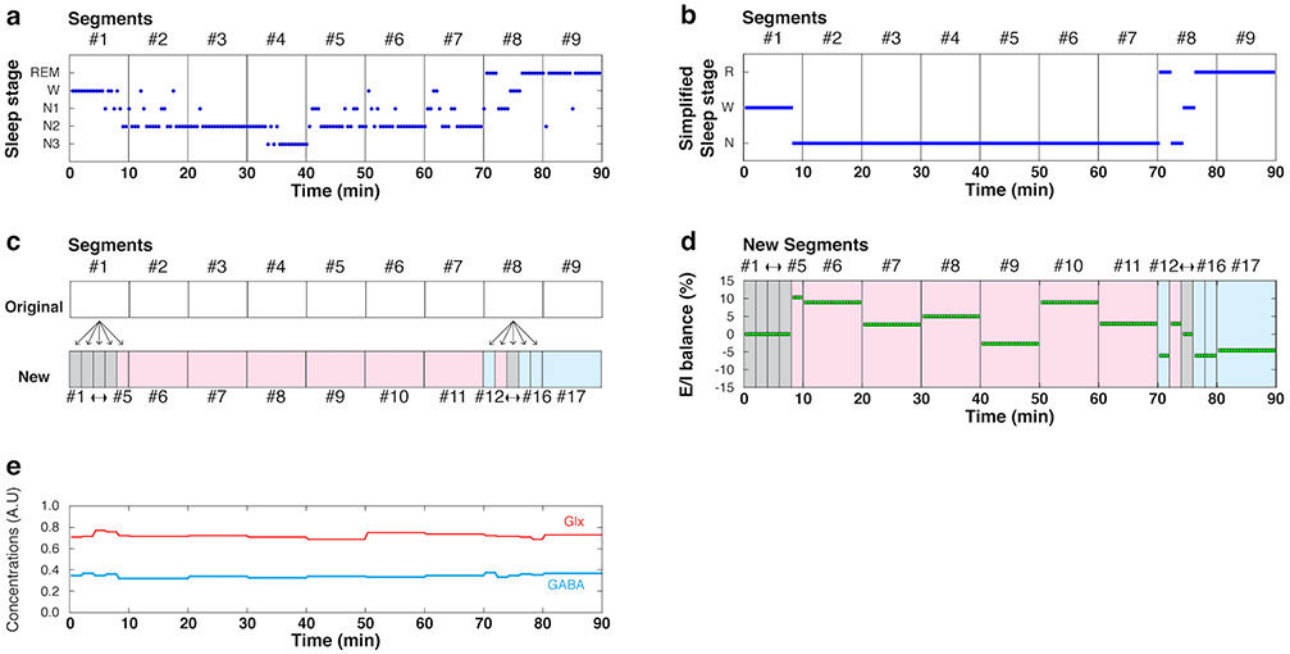
The value below each plot shows the mean full-width-at-half-maximum linewidth for NAA in Hz (mean ± SEM, Experiment 1, 8.0 ± 0.07 Hz; Experiment 3, 8.0 ± 0.05 Hz).



Extended Data Fig. 5. Correlation between EEG features, performance changes, and E/I balance changes during NREM and REM sleep.

We investigated whether sigma-band (13-16 Hz) and delta-band (1-4 Hz) activities during NREM sleep and theta-band (5-7 Hz) activity during REM sleep were involved in off-line performance gain or resilience to interference. We focused on these oscillatory activities because they are implicated in learning and memory[43-48]. First, a fast-Fourier transformation was applied to the EEG data in 5-sec epochs, and the data were smoothed with a tapered cosine window[44] to compute brain activities. Second, six epochs were used to yield the mean spectral for 30 s. Third, we calculated the power for each frequency band during both sleep and wakefulness in the trained region of early visual areas using a set of O1, PO3 and PO7 EEG channels or a set of O2, PO4, and PO8 EEG channels, depending on whether the target appeared in the right or left upper visual field. Fourth, we calculated the power for each frequency band during both sleep and wakefulness in the MT region, which was considered to be a control region according to a previous paper[44], using P7 and P8 EEG channels. Fifth, we normalized the power for each frequency band in both the trained region and the control region by subtracting the power for each frequency during sleep from that during wakefulness and dividing that value by that during wakefulness for each region to obtain the power of each frequency during sleep. Finally, we obtained the trained-region specific power during sleep for each frequency band by subtracting the normalized power in the control region from the normalized power in the trained region. Because NREM sleep was associated with off-line performance gains (see Fig. 1c & d), we measured the Pearson's correlation coefficients for sigma power during NREM sleep and off-line performance gain and for delta power during NREM sleep and off-line performance gain. Analogously, because REM sleep was associated with resilience to interference (see Fig. 1e & f), we measured the Pearson's correlation for theta-band power during REM sleep and resilience to interference. First, sigma power during NREM sleep was mildly correlated with off-line performance gains (**a**, n = 19 subjects, Pearson's r_{17} = 0.458, two-sided *t*-test, p = 0.048, 95% CI [0.01, 0.76], with a Bonferroni adjusted alpha level of 0.025 (0.05/2)) and E/I balance during NREM sleep (**c**, n = 19 subjects, Pearson's r_{17} = 0.397, two-sided *t*-test, p = 0.093, 95% CI [−0.070, 0.721]). Second, theta power during REM sleep was mildly correlated with resilience to interference (**b**, n = 10 subjects, Pearson's r_8=0.632, two-sided *t*-test, p = 0.050, with a Bonferroni adjusted alpha level of 0.025 (0.05/2), 95% CI [0.003, 0.902]) and E/I balance during REM sleep (**d**, n = 10 subjects, Pearson's r_8= −0.68, two-sided *t*-test, p = 0.031, 95% CI [−0.916, −0.084]). Finally, delta power during NREM sleep

was not significantly correlated with performance gains (n = 19 subjects, Pearson's r_{17} = 0.23, two-sided *t*-test, $p = 0.334$) or E/I balance (n = 19 subjects, Pearson's $r_{17} = 0.38$, two-sided *t*-test, $p = 0.113$) during NREM sleep. These results suggest that EEG power is only mildly correlated with off-line performance gains or with resistance to retrograde interference. However, the PSG data were obtained in a strong magnetic environment, which may not be suitable for performing a detailed power analysis.



Extended Data Fig. 6. An illustration of how to co-register MRS data and PSG data.

a, Sleep stage as a function of time (min). There are 9 segments for each 10-min MRS segment. In each segment, there are 20 epochs (each 30 sec) of sleep stage scoring. Each blue dot represents the sleep stage score for 30 sec. **b**, Simplified sleep stage as a function of time. The sleep stage score is simplified as 3 PSG states: Wake (W), NREM sleep (N; stage N1, N2 and N3 combined), and REM sleep (R). **c**, Splitting of two 10-min MRS segments (original #1 and #8) into five 2-min segments. This was because there were 2 or 3 PSG states within one original 10-min segment. In a "New" segments row, a gray section indicates Wake, a pink section indicates NREM sleep, and a blue section indicates REM sleep. **d**, The E/I balance during NREM sleep and REM sleep for each MRS segment was obtained by normalizing them to that during Wake. **e**, Time course of the concentrations of Glx (red) and GABA (cyan) underlying the E/I balance. Between the transition from **e** to **d**, there were several steps, as described in "Calculation of the mean E/I balance for NREM and REM sleep" in "Co-registration of MRS data and sleep stages" in Methods.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH (R21EY028329, R01EY019466, R01EY027841, T32EY018080, and T32MH115895) and BSF2016058. Part of this research was also supported by the Center for Vision Research, Brown University.

References

1. Rechtschaffen A & Kales A A Manual of Standardized Terminology, Techniques, and Scoring System for Sleep Stages of Human Subjects (Public Health Service, US Government Printing Office, Washington, DC, 1968).
2. Gais S, Plihal W, Wagner U & Born J Early sleep triggers memory for early visual discrimination skills. Nature neuroscience 3, 1335–1339 (2000). [PubMed: 11100156]
3. Tamaki M, et al. Enhanced spontaneous oscillations in the supplementary motor area are associated with sleep-dependent offline learning of finger-tapping motor-sequence task. J Neurosci 33, 13894–13902 (2013). [PubMed: 23966709]
4. Tononi G & Cirelli C Sleep and the price of plasticity: from synaptic and cellular homeostasis to memory consolidation and integration. Neuron 81, 12–34 (2014). [PubMed: 24411729]
5. Maquet P, et al. Experience-dependent changes in cerebral activation during human REM sleep. Nature neuroscience 3, 831–836 (2000). [PubMed: 10903578]
6. Karni A, Tanne D, Rubenstein BS, Askenasy JJ & Sagi D Dependence on REM sleep of overnight improvement of a perceptual skill. Science 265, 679–682 (1994). [PubMed: 8036518]
7. Diekelmann S & Born J The memory function of sleep. Nat Rev Neurosci 11, 114–126 (2010). [PubMed: 20046194]
8. Giuditta A, et al. The sequential hypothesis of the function of sleep. Behav Brain Res 69, 157–166 (1995). [PubMed: 7546307]
9. Stickgold R, James L & Hobson JA Visual discrimination learning requires sleep after training. Nature neuroscience 3, 1237–1238 (2000). [PubMed: 11100141]
10. Fischer S, Hallschmid M, Elsner AL & Born J Sleep forms memory for finger skills. Proc Natl Acad Sci U S A 99, 11987–11991 (2002). [PubMed: 12193650]
11. Ellenbogen JM, Hulbert JC, Stickgold R, Dinges DF & Thompson-Schill SL Interfering with theories of sleep and memory: sleep, declarative memory, and associative interference. Curr Biol 16, 1290–1294 (2006). [PubMed: 16824917]
12. Albouy G, et al. Cerebral Activity Associated with Transient Sleep-Facilitated Reduction in Motor Memory Vulnerability to Interference. Sci Rep 6, 34948 (2016). [PubMed: 27725727]
13. Sonni A & Spencer RM Sleep protects memories from interference in older adults. Neurobiol Aging 36, 2272–2281 (2015). [PubMed: 25890819]
14. Dudai Y The neurobiology of consolidations, or, how stable is the engram? Annu Rev Psychol 55, 51–86 (2004). [PubMed: 14744210]
15. Stickgold R Sleep-dependent memory consolidation. Nature 437, 1272–1278 (2005). [PubMed: 16251952]
16. Puentes-Mestril C & Aton SJ Linking Network Activity to Synaptic Plasticity during Sleep: Hypotheses and Recent Data. Front Neural Circuits 11, 61 (2017). [PubMed: 28932187]
17. Dosher BA & Lu ZL Hebbian Reweighting on Stable Representations in Perceptual Learning. Learn Percept 1, 37–58 (2009). [PubMed: 20305755]
18. Watanabe T, Nanez JE & Sasaki Y Perceptual learning without perception. Nature 413, 844–848 (2001). [PubMed: 11677607]
19. Hensch TK Critical period plasticity in local cortical circuits. Nat Rev Neurosci 6, 877–888 (2005). [PubMed: 16261181]
20. Bang JW, et al. Consolidation and reconsolidation share behavioral and neurochemical mechanisms. Nat Hum Behav 2, 507–513 (2018). [PubMed: 30505952]
21. Shibata K, et al. Overlearning hyperstabilizes a skill by rapidly making neurochemical processing inhibitory-dominant. Nature neuroscience 20, 470–475 (2017). [PubMed: 28135242]

22. Shadmehr R & Holcomb HH Neural correlates of motor memory consolidation. Science 277, 821–825 (1997). [PubMed: 9242612]
23. Seitz AR, et al. Task-specific disruption of perceptual learning. Proc Natl Acad Sci U S A 102, 14895–14900 (2005). [PubMed: 16203984]
24. Karni A & Sagi D Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. Proc Natl Acad Sci U S A 88, 4966–4970 (1991). [PubMed: 2052578]
25. Yotsumoto Y, Chang LH, Watanabe T & Sasaki Y Interference and feature specificity in visual perceptual learning. Vision Res 49, 2611–2623 (2009). [PubMed: 19665036]
26. Mednick SC, Cai DJ, Shuman T, Anagnostaras S & Wixted JT An opportunistic theory of cellular and systems consolidation. Trends Neurosci 34, 504–514 (2011). [PubMed: 21742389]
27. Rothman DL, Petroff OA, Behar KL & Mattson RH Localized 1H NMR measurements of gamma-aminobutyric acid in human brain in vivo. Proc Natl Acad Sci U S A 90, 5662–5666 (1993). [PubMed: 8516315]
28. Sanacora G, et al. Increased cortical GABA concentrations in depressed patients receiving ECT. Am J Psychiatry 160, 577–579 (2003). [PubMed: 12611844]
29. Barron HC, et al. Unmasking Latent Inhibitory Connections in Human Cortex to Reveal Dormant Cortical Memories. Neuron 90, 191–203 (2016). [PubMed: 26996082]
30. Stagg CJ, et al. Polarity-sensitive modulation of cortical neurotransmitters by transcranial stimulation. J Neurosci 29, 5202–5206 (2009). [PubMed: 19386916]
31. Stanley JA, et al. Functional dynamics of hippocampal glutamate during associative learning assessed with in vivo (1)H functional magnetic resonance spectroscopy. Neuroimage 153, 189–197 (2017). [PubMed: 28363835]
32. Floyer-Lea A, Wylezinska M, Kincses T & Matthews PM Rapid modulation of GABA concentration in human sensorimotor cortex during motor learning. J Neurophysiol 95, 1639–1644 (2006). [PubMed: 16221751]
33. Dinse HR, Ragert P, Pleger B, Schwenkreis P & Tegenthoff M Pharmacological modulation of perceptual learning and associated cortical reorganization. Science 301, 91–94 (2003). [PubMed: 12843392]
34. Diering GH, et al. Homer1a drives homeostatic scaling-down of excitatory synapses during sleep. Science 355, 511–515 (2017). [PubMed: 28154077]
35. Chung S, et al. Identification of preoptic sleep neurons using retrograde labelling and gene profiling. Nature 545, 477–481 (2017). [PubMed: 28514446]
36. Saper CB & Fuller PM Wake-sleep circuitry: an overview. Curr Opin Neurobiol 44, 186–192 (2017). [PubMed: 28577468]
37. Lee SH, et al. Activation of specific interneurons improves V1 feature selectivity and visual perception. Nature 488, 379–383 (2012). [PubMed: 22878719]
38. Li W, Ma L, Yang G & Gan WB REM sleep selectively prunes and maintains new synapses in development and learning. Nature neuroscience 20, 427–437 (2017). [PubMed: 28092659]
39. Wilson MA & McNaughton BL Reactivation of hippocampal ensemble memories during sleep. Science 265, 676–679 (1994). [PubMed: 8036517]
40. Riedel G, Platt B & Micheau J Glutamate receptor function in learning and memory. Behav Brain Res 140, 1–47 (2003). [PubMed: 12644276]
41. Niethard N, Burgalossi A & Born J Plasticity during Sleep Is Linked to Specific Regulation of Cortical Circuit Activity. Front Neural Circuits 11, 65 (2017). [PubMed: 28966578]
42. McDevitt EA, Duggan KA & Mednick SC REM sleep rescues learning from interference. Neurobiol Learn Mem 122, 51–62 (2015). [PubMed: 25498222]
43. Bang JW, Khalilzadeh O, Hamalainen M, Watanabe T & Sasaki Y Location specific sleep spindle activity in the early visual areas and perceptual learning. Vision Res 99, 162–171 (2014). [PubMed: 24380705]
44. Tamaki M, Wang Z, Watanabe T & Sasaki Y Trained-feature-specific offline learning by sleep in an orientation detection task. J Vis 19, 12 (2019).
45. Tamaki M, et al. Reward does not facilitate visual perceptual learning until sleep occurs. Proc Natl Acad Sci U S A 117, 959–968 (2020). [PubMed: 31892542]

46. Tononi G & Cirelli C Sleep and synaptic homeostasis: a hypothesis. Brain Res. Bull. 62, 143–150 (2003). [PubMed: 14638388]
47. Boyce R, Glasgow SD, Williams S & Adamantidis A Causal evidence for the role of REM sleep theta rhythm in contextual memory consolidation. Science 352, 812–816 (2016). [PubMed: 27174984]
48. Nishida M, Pearsall J, Buckner RL & Walker MP REM sleep, prefrontal theta, and the consolidation of human emotional memory. Cereb Cortex 19, 1158–1166 (2009). [PubMed: 18832332]
49. Faul F, Erdfelder E, Lang AG & Buchner A G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 39, 175–191 (2007). [PubMed: 17695343]
50. Tamaki M, Bang JW, Watanabe T & Sasaki Y The first-night effect suppresses the strength of slow-wave activity originating in the visual areas during sleep. Vision Res 99, 154–161 (2014). [PubMed: 24211789]
51. Tamaki M, Bang JW, Watanabe T & Sasaki Y Night watch in one brain hemisphere during sleep associated with the first-night effect in humans. Curr Biol 26, 1190–1194 (2016). [PubMed: 27112296]
52. Roenneberg T, Wirz-Justice A & Merrow M Life between clocks: daily temporal patterns of human chronotypes. J Biol Rhythms 18, 80–90 (2003). [PubMed: 12568247]
53. Green CS & Bavelier D Action video game modifies visual selective attention. Nature 423, 534–537 (2003). [PubMed: 12774121]
54. Berard AV, Cain MS, Watanabe T & Sasaki Y Frequent video game players resist perceptual interference. PLoS One 10, e0120011 (2015). [PubMed: 25807394]
55. Horne JA & Ostberg O A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. Int J Chronobiol 4, 97–110 (1976). [PubMed: 1027738]
56. Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR & Kupfer DJ The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 28, 193–213 (1989). [PubMed: 2748771]
57. Lubin A, Hord DJ, Tracy ML & Johnson LC Effects of exercise, bedrest and napping on performance decrement during 40 hours. Psychophysiology 13, 334–339 (1976). [PubMed: 951475]
58. Hoddes E, Vincent Z & Dement WC The development and use of the stanford sleepiness scale (SSS). Psychophysiology 9, 150 (1972).
59. Hoddes E, Zarcone V, Smythe H, Phillips R & Dement WC Quantification of sleepiness: a new approach. Psychophysiology 10, 431–436 (1973). [PubMed: 4719486]
60. Dinges DF & Powell JW Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. Behav Res Methods Instrum Comput 17, 652–655 (1985).
61. Tamaki M, Nittono H & Hori T The first-night effect occurs at the sleep-onset period regardless of the temporal anxiety level in healthy students. Sleep Biol. Rhythms 3, 92–94 (2005).
62. Agnew HW Jr., Webb WB & Williams RL The first night effect: an EEG study of sleep. Psychophysiology 2, 263–266 (1966). [PubMed: 5903579]
63. Carskadon MA & Dement WC Cumulative effects of sleep restriction on daytime sleepiness. Psychophysiology 18, 107–113 (1981). [PubMed: 6111825]
64. Schwartz S, Maquet P & Frith C Neural correlates of perceptual learning: a functional MRI study of visual texture discrimination. Proc Natl Acad Sci U S A 99, 17137–17142 (2002). [PubMed: 12446842]
65. Yotsumoto Y, et al. Location-specific cortical activation changes during sleep after training for perceptual learning. Curr Biol 19, 1278–1282 (2009). [PubMed: 19576772]
66. Brainard DH The Psychophysics Toolbox. Spatial vision 10, 433–436 (1997). [PubMed: 9176952]
67. Pelli DG The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spatial vision 10, 437–442 (1997). [PubMed: 9176953]

68. Machizawa M, Patey R, Kim D & Watanabe T Different aspects of training on a texture discrimination task (TDT) improves different attentional abilities. J Vis 14, 951 (2014). DOI: 10.1167/14.10.951
69. Wichmann FA & Hill NJ The psychometric function: I. Fitting, sampling, and goodness of fit. Percept Psychophys 63, 1293–1313 (2001). [PubMed: 11800458]
70. Luce RD Response time distributions in memory search: a caution in Human Memory and Cognitive Capabilities (ed. Klix F & Hagendorf H) 109–121 (Elsevier Science Publishers B.V, North-Holland, 1986).
71. Lim J & Dinges DF Sleep deprivation and vigilant attention. Ann N Y Acad Sci 1129, 305–322 (2008). [PubMed: 18591490]
72. Iber C, Ancoli-Israel S, Chesson A & Quan SF The AASM manual for the scoring of sleep and associated events: rules, terminology, and technical specification. (American Academy of Sleep Medicine, Westchester, IL, 2007).
73. Sharbrough FCGE, Lesser RP, Lüders H, Nuwer M & Picton TW American electroencephalographic society guidelines for standard electrode position nomenclature. J. Clin. Neurophysiol. 8, 200–202 (1991). [PubMed: 2050819]
74. Mescher M, Merkle H, Kirsch J, Garwood M & Gruetter R Simultaneous in vivo spectral editing and water suppression. NMR Biomed 11, 266–272 (1998). [PubMed: 9802468]
75. Edden RA & Barker PB Spatial effects in the detection of gamma-aminobutyric acid: improved sensitivity at high fields using inner volume saturation. Magn Reson Med 58, 1276–1282 (2007). [PubMed: 17969062]
76. Hu Y, Chen X, Gu H & Yang Y Resting-state glutamate and GABA concentrations predict task-induced deactivation in the default mode network. J Neurosci 33, 18566–18573 (2013). [PubMed: 24259578]
77. Muthukumaraswamy SD, Edden RA, Jones DK, Swettenham JB & Singh KD Resting GABA concentration predicts peak gamma frequency and fMRI amplitude in response to visual stimulation in humans. Proc Natl Acad Sci U S A 106, 8356–8361 (2009). [PubMed: 19416820]
78. Stagg CJ, Bachtiar V & Johansen-Berg H The role of GABA in human motor learning. Curr Biol 21, 480–484 (2011). [PubMed: 21376596]
79. Robertson CE, Ratai EM & Kanwisher N Reduced GABAergic Action in the Autistic Brain. Curr Biol 26, 80–85 (2016). [PubMed: 26711497]
80. Henry ME, Lauriat TL, Shanahan M, Renshaw PF & Jensen JE Accuracy and stability of measuring GABA, glutamate, and glutamine by proton magnetic resonance spectroscopy: a phantom study at 4 Tesla. J Magn Reson 208, 210–218 (2011). [PubMed: 21130670]
81. Stagg CJ, et al. Neurochemical effects of theta burst stimulation as assessed by magnetic resonance spectroscopy. J Neurophysiol 101, 2872–2877 (2009). [PubMed: 19339458]
82. Stagg CJ, et al. Local GABA concentration is related to network-level resting functional connectivity. Elife 3, e01465 (2014). [PubMed: 24668166]
83. Tkác I, Starcuk Z, Choi IY & Gruetter R In vivo 1H NMR spectroscopy of rat brain at 1 ms echo tim. Magn. Reson. Med. 41, 649–656 (1999). [PubMed: 10332839]
84. Rothman DL, Behar KL, Hetherington HP & Shulman RG Homonuclear 1H double-resonance difference spectroscopy of the rat brain in vivo. Proc Natl Acad Sci U S A 81, 6330–6334 (1984). [PubMed: 6149543]
85. Klose U In vivo proton spectroscopy in presence of eddy currents. Magn Reson Med 14, 26–30 (1990). [PubMed: 2161984]
86. Oeltzschner G, Schnitzler A, Wickrath F, Zollner HJ & Wittsack HJ Use of quantitative brain water imaging as concentration reference for J-edited MR spectroscopy of GABA. Magn Reson Imaging 34, 1057–1063 (2016). [PubMed: 27109486]
87. Gasparovic C, et al. Use of tissue water as a concentration reference for proton spectroscopic imaging. Magn Reson Med 55, 1219–1226 (2006). [PubMed: 16688703]
88. Provencher SW Estimation of metabolite concentrations from localized in vivo proton NMR spectra. Magn Reson Med 30, 672–679 (1993). [PubMed: 8139448]
89. Provencher SW Automatic quantitation of localized in vivo 1H spectra with LCModel. NMR Biomed 14, 260–264 (2001). [PubMed: 11410943]

90. Lunghi C, Emir UE, Morrone MC & Bridge H Short-term monocular deprivation alters GABA in the adult human visual cortex. Curr Biol 25, 1496–1501 (2015). [PubMed: 26004760]

91. Zeinali-Rafsanjani B, et al. MRS Shimming: An Important Point Which Should not be Ignored. J Biomed Phys Eng 8, 261–270 (2018). [PubMed: 30320030]

92. Kreis R Issues of spectral quality in clinical 1H-magnetic resonance spectroscopy and a gallery of artifacts. NMR Biomed 17, 361–381 (2004). [PubMed: 15468083]

93. Cohen J Statistical Power Analysis for the Behavioral Sciences (Lawrence Erlbaum and Associates, Hillsdale, 1988).

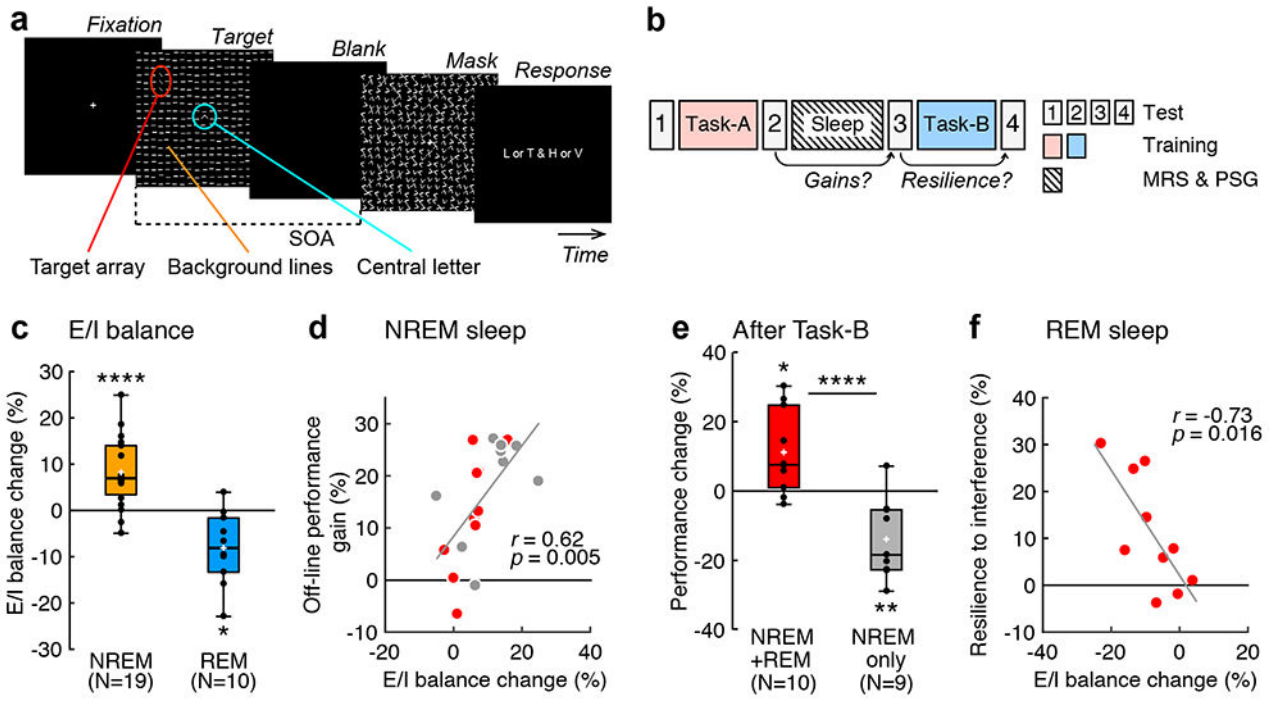


Fig. 1.

Procedures and results of Experiment 1. **a**, A TDT trial. See Texture discrimination task (TDT) in Methods for more details. **b**, Experimental design. Subjects were trained on the first task (Task-A) before a nap and on the second task (Task-B) after the nap. **c**, Boxplots for the E/I balance changes from baseline to NREM sleep (orange, n = 19 subjects, **** two-sided one sample *t*-test against 0, $t_{18} = 4.77$, $p < 0.001$, Cohen's $d = 1.09$, 95% CI [4.61, 11.88]) and REM sleep (blue, n = 10 subjects, * two-sided one sample *t*-test against 0, $t_9 = 3.21$, $p = 0.011$, Cohen's $d = 1.02$, 95% CI [−13.82, −2.40]). **d**, Scatter plots (n = 19 subjects, Pearson's $r_{17} = 0.62$, two-sided *t*-test, $t_{17} = 3.25$, $p = 0.005$, 95% CI [0.23, 0.84]) of E/I balance changes from baseline to NREM sleep (the horizontal axis) vs. off-line performance gain (the vertical axis indicates the performance changes from the second to the third test sessions on Task-A) including the NREM+REM group (red circles, n = 10 subjects) and the NREM-only group (gray circles, n = 9 subjects). **e**, Boxplots for the degree of resilience to retrograde interference (performance changes from the third to fourth test sessions for Task-A) for the NREM+REM (red, n = 10 subjects, * two-sided one sample *t*-test against 0, $t_9 = 2.89$, $p = 0.019$, Cohen's $d = 0.91$, 95% CI [2.44, 19.93]) and NREM-only (gray, n = 9 subjects, ** two-sided one sample *t*-test against 0, $t_8 = 3.60$, $p = 0.007$, Cohen's $d = 1.20$, 95% CI [−22.74, −4.98]) groups. **** two-sided independent-samples *t*-test between groups, n = 19 subjects, $t_{17} = 4.58$, $p < 0.001$, Cohen's $d = 2.10$, 95% CI [5.58, 13.50]. **f**, Scatter plots (n = 10 subjects, Pearson's $r_8 = -0.73$, two-sided *t*-test, $t_8 = 3.05$, $p = 0.016$, 95% CI [−0.93, −0.19) of the E/I balance during REM sleep (the horizontal axis) vs. the mean degree of resilience to retrograde interference (the vertical axis) for the NREM +REM group (red circles). In each boxplot, the bottom and the top of the box correspond to the 25th and 75th percentiles (the lower and upper quartiles), respectively. The inner thick horizontal line represents the median, and the plus mark represents the mean. The whiskers

show the maximum and minimum of the data. Individual data (dots) are overlaid. Grubbs' test showed no outliers (Alpha=.05, two-sided).

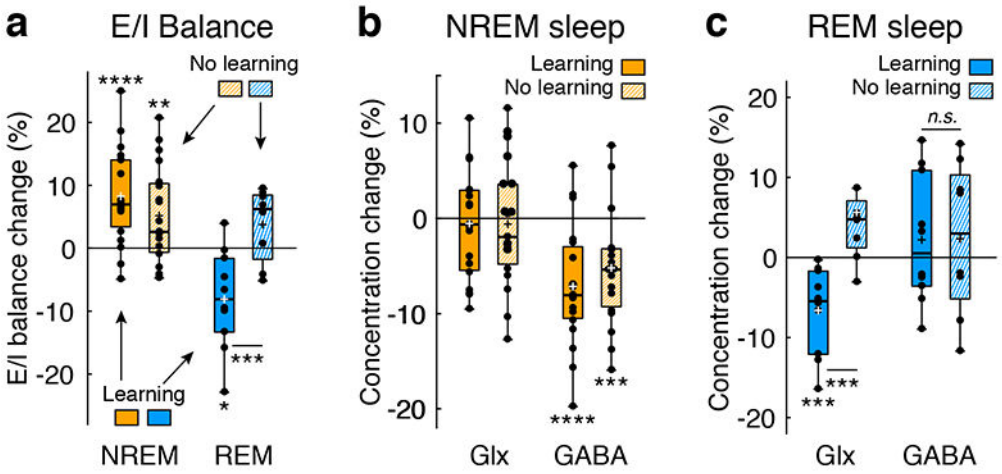


Fig. 2.

Collective results of Experiments 1 and 3. Filled orange and blue bars represent NREM sleep and REM sleep, respectively, in Experiment 1 (with learning). Shaded orange and blue bars represent NREM sleep and REM sleep, respectively, in Experiment 3 (with no learning). **a**, Boxplots for the E/I balance during NREM and REM sleep after learning (n = 19 subjects for NREM sleep, n = 10 subjects for REM sleep, the same as Fig. 1c) in Experiment 1 and after no prior learning (the control condition, n = 19 subjects for NREM sleep, ** two-sided one sample *t*-test against 0, $t_{18} = 2.91$, $p = 0.009$, Cohen's $d = 0.67$, 95% CI [1.44, 8.87]; n = 8 subjects for REM sleep, two-sided one sample *t*-test against 0, $t_7 = 1.80$, $p = 0.115$) in Experiment 3. The E/I balance during REM sleep was significantly different between Exps. 1 (n = 10 subjects) and 3 (n = 8 subjects), *** two-sided independent-samples *t*-test between Experiments, n = 18 subjects, $t_{16} = 3.50$, $p = 0.003$, Cohen's $d = 1.66$, 95% CI [4.68, 19.04].**b**, Boxplots for the concentrations of Glx (n = 19 subjects for Exp. 1, two-sided one sample *t*-test against 0, $t_{18} = 0.40$, $p = 0.691$; n = 19 subjects for Exp. 3, two-sided one-sample *t*-test against 0, $t_{18} = 0.40$, $p = 0.695$) and GABA (n = 19 subjects for Exp. 1, **** two-sided one sample *t*-tests against 0, $t_{18} = 4.93$, $p < 0.001$, Cohen's $d = 1.13$, 95% CI [−10.18, −4.10]; n = 19 subjects for Exp. 3, *** two-sided one sample *t*-tests against 0, $t_{18} = 3.76$, $p = 0.001$, Cohen's $d = 0.86$, 95% CI [−8.11, −2.30]) during NREM sleep. **c**, Boxplots for the concentrations of Glx (n = 10 subjects for Exp. 1, *** two-sided one sample *t*-test against 0, $t_9 = 3.87$, $p = 0.004$, Cohen's $d = 1.23$, 95% CI [−10.57, −2.78]; n = 8 subjects for Exp. 3, one sample *t*-test against 0, $t_7 = 2.09$, $p = 0.075$; *** two-sided independent-samples *t*-test between Experiments, n = 18 subjects, $t_{16} = 4.00$, $p = 0.001$, Cohen's $d = 1.90$, 95% CI [−18.74, −5.77]) and GABA (n = 10 subjects for Exp. 1, n = 8 subjects for Exp. 3; two-sided independent-samples *t*-test between Experiments, n = 18 subjects, $t_{16} = 0.03$, $p = 0.975$) during REM sleep. For each boxplot, the bottom and top of the box correspond to the 25th and 75th percentiles (the lower and upper quartiles), respectively. The inner thick horizontal line represents the median, and the plus mark represents the mean. The whiskers show the maximum and minimum of the data. Individual data (dots) are overlaid. Grubbs' test showed no outliers (Alpha=.05, two-sided). In each plot, zero represents baseline, which was measured during wakefulness.

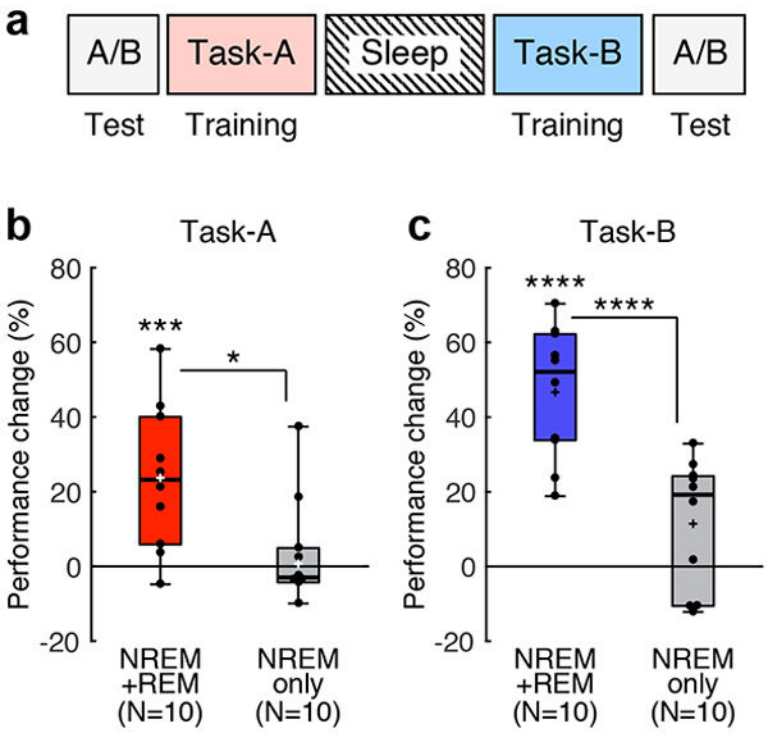


Fig. 3.

The experimental design and results of Experiment 4. **a**, Design. The first TDT training was conducted with background A (Task-A, pink), and the second training was conducted with background B (Task-B, cyan). Test sessions were conducted before the first training (pretest) and after the interval and the second training (posttest) to examine performance changes on both Task-A and Task-B. **b**, Boxplots for the performance change for Task-A from the pretest to the posttest sessions for NREM+REM group (red, n = 10 subjects, *** two-sided one sample *t*-test against 0, $t_9 = 3.83$, $p = 0.004$, Cohen's $d = 1.21$, 95% CI [9.68, 37.69]) and NREM-only group (gray, n = 10 subjects, two-sided one sample *t*-test against 0, $t_9 = 0.12$, $p = 0.904$). * post hoc two-sided independent-samples *t*-test between groups, n = 20 subjects, $t_{18} = 2.72$, $p = 0.014$, Cohen's $d = 1.22$, 95% CI [5.24, 40.71]. **c**, Boxplots for the performance change for Task-B from the pretest to the posttest sessions for NREM+REM group (blue, n = 10 subjects, **** two-sided one sample *t*-test against 0, $t_9 = 8.27$, $p < 0.001$, Cohen's $d = 2.62$, 95% CI [33.90, 59.43]) and NREM-only group (gray, n = 10 subjects, two-sided one sample *t*-test against 0, $t_9 = 2.06$, $p = 0.069$). **** post hoc two-sided independent-samples *t*-test between groups, n = 20 subjects, $t_{18} = 4.45$, $p < 0.001$, Cohen's $d = 1.99$, 95% CI [18.59, 51.84]. For each boxplot, the bottom and top of the box correspond to the 25th and 75th percentiles (the lower and upper quartiles), respectively. The inner thick horizontal line represents the median, and the plus mark represents the mean. The whiskers show the maximum and minimum of the data. Individual data (dots) are overlaid. Grubbs' test showed no outliers (Alpha=.05, two-sided).

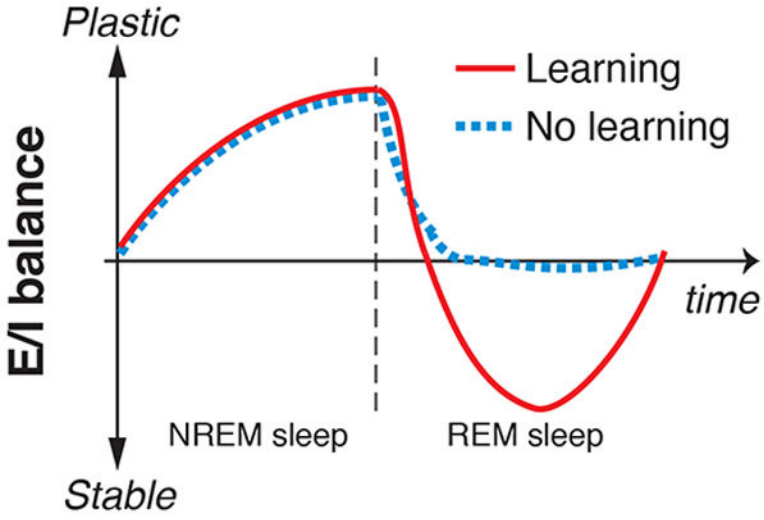


Fig. 4.

An illustration of the complementary process hypothesis. Plasticity/stability modulation (E/I balance) during NREM and REM sleep with or without presleep learning is plotted as a function of time.

Table 1.

Functional and neurochemical differences between NREM and REM sleep

	NREM sleep	REM sleep
E/I balance (plasticity)	Up	Down
Performance gain	Yes	No
Stabilization	No	Yes
Learning specificity	No	Yes