**Supplementary Table 1. Subject information.**

Twenty subjects from Exp. 2 also participated in Exp. 4, where performances for both Tasks A and B were measured.

| Experiment | The number of subjects analyzed | The number of females | Age (mean ± SEM) | Data omission | The number of recruitments |
|---|---|---|---|---|---|
| 1 | 19 | 15 | 23.7 ± 0.66 | 1 | 20 |
| 2 | 38 | 26 | 22.6 ± 0.52 | 3 | 41 |
| 3 | 19 | 8 | 24.2 ± 0.88 | 1 | 20 |
| 4 | 20 | 17 | 22.8 ± 0.63 | 2 | 22 |

**Supplementary Table 2.** Sleep parameters for Experiments 1-4. We tested whether the sleep quality during Experiments 1 and 3 was different, as we directly compared E/I balances during sleep between them as discussed in the main text. Because the Shapiro-Wilk test showed a violation of normality for most of the parameters, we used Kruskal-Wallis one-way ANOVA for the comparison between Experiments 1 and 3. The statistical results indicated no significant difference in any of the sleep parameters between Experiments 1 and 3. All sleep parameters except TIB were obtained from the first sleep cycle. Note that data from the NREM+REM group only were used for REM (%) and REM (min), as the NREM-only group did not have REM sleep. Otherwise, data from both groups were combined.

| | **Exp. 1 (n=19 subjects)** | **Exp. 2 (n=38 subjects)** | **Exp. 3 (n=19 subjects)** | **Exp. 4 (n=20 subjects)** | **Kruskal-Wallis (df=2)** | |
|---|---|---|---|---|---|---|
| | | | | | **Chi square** | **P value** |
| SOL (min) | 6.7 ± 1.15 | 8.8 ± 1.20 | 5.6 ± 1.59 | 8.2 ± 1.88 | 2.06 | 0.152 |
| WASO (min) | 5.6 ± 1.86 | 4.6 ± 1.42 | 12.0 ± 2.78 | 5.1 ± 2.34 | 2.54 | 0.111 |
| Stage W (%) | 16.4 ± 2.84 | 14.2 ± 2.49 | 23.1 ± 4.24 | 15.5 ± 4.12 | 1.26 | 0.261 |
| NREM (%) | 78.1 ± 3.13 | 78.7 ± 2.30 | 72.9 ± 3.81 | 79.4 ± 3.82 | 1.26 | 0.261 |
| REM (%) | 10.5 ± 2.03 | 11.7 ± 1.47 | 9.4 ± 1.99 | 10.3 ± 2.23 | 0.24 | 0.625 |
| Stage W (min) | 9.6 ± 1.93 | 10.0 ± 1.70 | 15.3 ± 2.73 | 10.3 ± 2.73 | 2.68 | 0.102 |
| NREM (min) | 43.7 ± 3.25 | 56.2 ± 2.44 | 48.7 ± 3.75 | 51.3 ± 2.86 | 1.77 | 0.184 |
| REM (min) | 5.8 ± 1.30 | 9.3 ± 1.26 | 6.4 ± 1.50 | 8.2 ± 2.06 | 0.05 | 0.824 |
| SE (%) | 83.6 ± 2.84 | 85.7 ± 2.48 | 76.9 ± 4.24 | 84.3 ± 4.11 | 1.26 | 0.261 |
| TIB (min) | 71.2 ± 5.08 | 79.9 ± 2.74 | 80.2 ± 3.81 | 71.3 ± 4.28 | 1.51 | 0.219 |

**SOL**, sleep-onset latency. **WASO**, wake after sleep onset. **SE**, sleep efficiency. **TIB**, time in bed, which indicates the duration of each sleep session (the time interval between lights-off and lights-on). NREM sleep includes NREM sleep stages 1-3. Values are the mean ± SEM. The P values were not corrected for multiple comparisons.

**Supplementary Table 3. Sleepiness data for Experiment 1.**

| **Experiment 1** | | | | | |
|---|---|---|---|---|---|
| Group | | 1st test session | 2nd test session | 3rd test session | 4th test session |
| NREM+REM (n=10 subjects) | SSS score | 1.5 ± 0.17 | 1.7 ± 0.15 | 1.2 ± 0.13 | 1.2 ± 0.13 |
| | Mean RT | 2.55 ± 0.01 | 2.54 ± 0.01 | 2.53 ± 0.01 | 2.54 ± 0.01 |
| | # of lapses | 0.8 ± 0.42 | 0.3 ± 0.21 | 0.2 ± 0.13 | 0.4 ± 0.22 |
| | 10% fastest RTs | 2.47 ± 0.01 | 2.48 ± 0.01 | 2.47 ± 0.01 | 2.48 ± 0.01 |
| | 10% slowest RTs | 2.63 ± 0.03 | 2.61 ± 0.02 | 2.58 ± 0.01 | 2.60 ± 0.02 |
| | Lapse threshold | 96.3% | 99.5% | 99.9% | 99.4% |
| NREM only (n=9 subjects) | SSS score | 1.8 ± 0.15 | 1.4 ± 0.18 | 1.1 ± 0.11 | 1.2 ± 0.15 |
| | Mean RT | 2.56 ± 0.02 | 2.56 ± 0.01 | 2.54 ± 0.01 | 2.54 ± 0.01 |
| | # of lapses | 0.9 ± 0.48 | 0.8 ± 0.36 | 0.2 ± 0.15 | 0.6 ± 0.24 |
| | 10% fastest RTs | 2.49 ± 0.01 | 2.49 ± 0.01 | 2.48 ± 0.01 | 2.48 ± 0.01 |
| | 10% slowest RTs | 2.65 ± 0.04 | 2.63 ± 0.02 | 2.61 ± 0.02 | 2.64 ± 0.04 |
| | Lapse threshold | 93.5% | 98.4% | 99.6% | 99.2% |

The results showed no significant difference in the SSS score (two-sided Mann-Whitney U test: 1st test session, $U = 32.5$, $p = 0.242$; 2nd test session, $U = 33.5$, $p = 0.294$; 3rd test session, $U = 41$, $p = 0.651$; 4th test session, $U = 44$, $p = 0.954$), mean RT (mixed-design ANOVA: Group, $F(1, 17) = 0.75$, $p = 0.398$; Test session, $F(3, 51) = 2.93$, $p = 0.042$; Group x Test session, $F(3, 51) = 0.49$, $p = 0.691$), number of lapses (two-sided Mann-Whitney U test: 1st test session, $U = 44.5$, $p = 1.000$; 2nd test session, $U = 33.5$, $p = 0.275$; 3rd test session, $U = 44$, $p = 0.954$; 4th test session, $U = 39$, $p = 0.599$), 10% fastest RTs (two-sided Mann-Whitney U test: 1st test session, $U = 24$, $p = 0.094$; 2nd test session, U = 39, $p = 0.653$; 3rd test session, $U = 33.5$, $p = 0.369$; 4th test session, $U = 39$, $p = 0.653$), or 10% slowest RTs (two-sided Mann-Whitney U test: 1st test session, $U = 41$, $p = 0.775$; 2nd test session, $U = 34$, $p = 0.391$; 3rd test session, $U = 34$, $p = 0.391$; 4th test session, $U = 36$, $p = 0.488$) between the groups in any of the test sessions. The *P* values were not corrected for multiple comparisons.

**Supplementary Table 4. Sleepiness data for Experiment 2.**

| **Experiment 2** | | | |
|---|---|---|---|
| Group | | Pre | Post |
| NREM+REM (n=23 subjects) | SSS score | 1.8 ± 0.08 | 1.4 ± 0.11 |
| | Mean RT | 2.51 ± 0.01 | 2.51 ± 0.01 |
| | # of lapses | 0.35 ± 0.14 | 0.48 ± 0.16 |
| | 10% fastest RTs | 2.40 ± 0.02 | 2.39 ± 0.03 |
| | 10% slowest RTs | 2.59 ± 0.01 | 2.59 ± 0.02 |
| | Lapse threshold | 99.7% | 98.2% |
| NREM only (n=15 subjects) | SSS score | 1.5 ± 0.13 | 1.5 ± 0.17 |
| | Mean RT | 2.51 ± 0.01 | 2.49 ± 0.01 |
| | # of lapses | 0.40 ± 0.16 | 0.47 ± 0.24 |
| | 10% fastest RTs | 2.41 ± 0.02 | 2.43 ± 0.01 |
| | 10% slowest RTs | 2.59 ± 0.01 | 2.58 ± 0.01 |
| | Lapse threshold | 99.9% | 99.9% |

The results showed no significant difference in the SSS score (two-sided Mann-Whitney U test: Pre, $U = 122$, $p = 0.057$; Post, $U = 162$, $p = 0.731$), mean RT (two-sided Mann-Whitney U test: Pre, $U = 171$, $p = 0.976$; Post, $U = 129$, $p = 0.199$), number of lapses (two-sided Mann-Whitney U test: Pre, $U = 162$, $p = 0.707$; Post, $U = 162.5$, $p = 0.729$), 10% fastest RTs (two-sided Mann-Whitney U test: Pre, $U = 157$, $p = 0.654$; Post, $U = 164$, $p = 0.811$), or 10% slowest RTs (two-sided Mann-Whitney U test: Pre, $U = 154.5$, $p = 0.601$; Post, $U = 161$, $p = 0.743$) between the groups in any of the test sessions.

**Supplementary Table 5. Sleepiness data for Experiment 3.**

SSS scores were measured in 14 subjects, and the PVT was not conducted in Exp. 3.

| **Experiment 3** | | | |
|---|---|---|---|
| Condition | | Presleep | |
| Control (n=14 subjects) | SSS score | 1.7 | ± 0.19 |

**Supplementary Table 6. Sleepiness data for Experiment 4.**

| **Experiment 4** | | | |
|---|---|---|---|
| Group | | Pre | Post |
| NREM+REM (n=10 subjects) | SSS score | 1.9 ± 0.10 | 1.6 ± 0.16 |
| | Mean RT | 2.52 ± 0.01 | 2.51 ± 0.02 |
| | # of lapses | 0.5 ± 0.22 | 0.7 ± 0.34 |
| | 10% fastest RTs | 2.36 ± 0.03 | 2.38 ± 0.04 |
| | 10% slowest RTs | 2.61 ± 0.02 | 2.61 ± 0.05 |
| | Lapse threshold | 99.4% | 94.7% |
| NREM only (n=10 subjects) | SSS score | 1.5 ± 0.17 | 1.3 ± 0.15 |
| | Mean RT | 2.52 ± 0.01 | 2.51 ± 0.02 |
| | # of lapses | 0.6 ± 0.22 | 0.6 ± 0.34 |
| | 10% fastest RTs | 2.43 ± 0.02 | 2.44 ± 0.01 |
| | 10% slowest RTs | 2.61 ± 0.02 | 2.59 ± 0.02 |
| | Lapse threshold | 99.5% | 99.7% |

The results showed no significant difference in the SSS score (two-sided Mann-Whitney U test: Pre, $U = 30$, $p = 0.0636$; Post, $U = 35$, $p = 0.204$), mean RT (mixed-design ANOVA: Group, $F(1, 18) = 0.01$, $p = 0.977$; Test session, $F(1, 10) = 0.15$, $p = 0.704$; Group x Test session, $F(1, 18) = 0.006$, $p = 0.940$), number of lapses (two-sided Mann-Whitney U test: Pre, $U = 45.5$, $p = 0.734$; Post, $U = 46$, $p = 0.756$), 10% fastest RTs (two-sided Mann-Whitney U test: Pre, $U = 23$, $p = 0.045$; Post, $U = 42$, $p = 0.571$), or 10% slowest RTs (two-sided Mann-Whitney U test: Pre, $U = 50$, $p = 1.000$; Post, $U = 49$, $p = 0.970$) between the groups in any of the test sessions. The P values were not corrected for multiple comparisons. Note that the results for "10% fastest RTs" showed a significant difference between the groups. However, there were no other results that showed a significant difference between the groups. Thus, it is not reasonable to assume that there was a significant difference in sleepiness between the groups.

**Supplementary Table 7.** Initial performance. Threshold SOA (ms) for the TDT. Because normality of the data was not assumed, we used a two-sided Mann-Whitney U test.

| | Group | | Threshold (ms) (Mean ± SEM) | Mann-Whitney U test (NREM+REM vs. NREM only) |
|---|---|---|---|---|
| Exp. 1 | NREM+REM (n=10 subjects) | | 142.0 ± 11.15 | $U = 43$, $p = 0.903$ |
| | NREM only (n=9 subjects) | | 151.2 ± 20.30 | |
| Exp. 2 | NREM+REM (n=23 subjects) | | 136.4 ± 9.70 | $U = 149$, $p = 0.492$ |
| | NREM only (n=15 subjects) | | 134.2 ± 16.45 | |
| Exp. 4 | NREM+REM (n=10 subjects) | Task-A | 139.0 ± 18.53 | Task A: $U = 47$, $p = 0.850$ |
| | | Task-B | 181.8 ± 20.50 | |
| | NREM only (n=10 subjects) | Task-A | 136.3 ± 22.90 | Task B: $U = 24$, $p = 0.054$ |
| | | Task-B | 134.5 ± 18.07 | |

**Supplementary Table 8.** Pearson's correlation coefficients for sleep duration time (min) and the 4 parameters in Exp. 1.The summation of NREM sleep and REM sleep (min) were not correlated with any performance change or E/I balance change in the present study. The significance of correlation coefficients was determined by two-sided *t* tests (not corrected for multiple comparisons). See also Justification for the sleep termination method in the **Methods** for more information.

| | | Off-line performance gains | Resilience to retrograde interference | E/I balance during NREM sleep | E/I balance during REM sleep |
|---|---|---|---|---|---|
| Sleep time (NREM+REM) | Correlation (r) | -0.043 | -0.044 | 0.198 | 0.212 |
| | P value | 0.861 | 0.857 | 0.416 | 0.556 |
| | N (subjects) | 19 | 19 | 19 | 10 |

**Supplementary Table 9.** Pearson's correlation coefficients for behavioral changes and sleep stage duration (min) in Exp. 1. The duration of NREM sleep and REM sleep were not correlated with any performance change in the present study. The significance of correlation coefficients was determined by two-sided *t* tests (not corrected for multiple comparisons). See also Justification for the sleep termination method in the **Methods** for more information.

| | | Duration of NREM sleep | Duration of REM sleep |
|---|---|---|---|
| Off-line performance gains | Correlation (r) | 0.020 | -0.052 |
| | P value | 0.937 | 0.888 |
| | N (subjects) | 19 | 10 |
| Resilience to retrograde interference | Correlation (r) | -0.194 | -0.068 |
| | P value | 0.426 | 0.852 |
| | N (subjects) | 19 | 10 |

**Supplementary Table 10.** Pearson's correlation coefficient matrix for the E/I balance during each sleep stage and the duration of NREM sleep, REM sleep and sleep time, using data from both Exps. 1 and 3. The duration of NREM sleep, REM sleep and the summation of NREM sleep and REM sleep (min) were not correlated with E/I balance change in the present study. The significance of correlation coefficients was determined by two-sided *t* tests (not corrected for multiple comparisons). See also Justification for the sleep termination method in the **Methods** for more information.

| | | E/I balance during NREM sleep | E/I balance during REM sleep |
|---|---|---|---|
| Duration of NREM sleep | Correlation (r) | 0.175 | 0.311 |
| | P value | 0.293 | 0.209 |
| | N (subjects) | 38 | 18 |
| Duration of REM sleep | Correlation (r) | 0.103 | 0.029 |
| | P value | 0.686 | 0.909 |
| | N (subjects) | 18 | 18 |
| Sleep time (NREM + REM) | Correlation (r) | 0.170 | 0.280 |
| | P value | 0.307 | 0.260 |
| | N (subjects) | 38 | 18 |

**Supplementary Table 11.** MRS data quality. Values are mean ± SEM.

| | Shim value (Hz) | NAA linewidth (Hz) | Frequency drift (Hz) | | %SD for Glx | | %SD for GABA | |
|---|---|---|---|---|---|---|---|---|
| | | | First run | Last run | NREM | REM | NREM | REM |
| **Experiment 1** (n = 19 subjects) | 13.8 ± 0.15 | 8.0 ± 0.07 | 1.3 ± 0.31 | 1.1 ± 0.21 | 6.6 ± 0.25 | 8.4 ± 0.79 | 8.3 ± 0.66 | 9.0 ± 1.05 |
| **Experiment 3** (n = 19 subjects) | 14.0 ± 0.17 | 8.0 ± 0.05 | 1.1 ± 0.14 | 1.8 ± 0.57 | 6.0 ± 0.21 | 6.7 ± 0.39 | 7.9 ± 0.28 | 9.0 ± 0.43 |

**Supplementary Table 12.** Morningness-Eveningness questionnaire (MEQ) and the Pittsburgh Sleep Quality Index.

| | | MEQ score | | | Global PSQI score | | | NREM+REM vs. NREM only |
|---|---|---|---|---|---|---|---|---|
| Exp. 1 (N=19 subjects) | NREM+REM | 57.4 | ± | 1.82 | 2.2 | ± | 0.39 | MEQ: $t$ (17) = 0.37, $p$ = 0.714<br>PSQI: $U$ = 33, $p$ = 0.316 |
| | NREM only | 56.4 | ± | 1.79 | 2.3 | ± | 0.37 | |
| Exp. 2 (N=38 subjects) | NREM+REM | 55.3 | ± | 1.56 | 2.6 | ± | 0.29 | MEQ: $t$ (36) = 0.17, $p$ = 0.862<br>PSQI: $U$ = 157, $p$ = 0.646 |
| | NREM only | 55.7 | ± | 1.58 | 2.8 | ± | 0.34 | |
| Exp. 3 (N=19 subjects) | NREM+REM | 58.5 | ± | 2.75 | 2.9 | ± | 0.40 | MEQ: $t$ (17) = 0.48, $p$ = 0.635<br>PSQI: $U$ = 31, $p$ = 0.285 |
| | NREM only | 56.8 | ± | 2.20 | 2.3 | ± | 0.43 | |
| Exp. 4 (N=20 subjects) | NREM+REM | 57.1 | ± | 2.80 | 2.7 | ± | 0.52 | MEQ: $t$ (18) = 0.34, $p$ = 0.737<br>PSQI: $U$ = 49.5, $p$ = 1.000 |
| | NREM only | 55.9 | ± | 2.13 | 2.8 | ± | 0.36 | |

***Note.*** Values are the mean ± SEM. The MEQ score was obtained from the Horne-Östberg's Morningness-Eveningness questionnaire[55]. The global PSQI score was obtained from the Pittsburgh Sleep Quality Index[56].

The MEQ and PSQI scores were compared between the NREM+REM and NREM-only groups. Since the MEQ scores were normally distributed, a two-sided independent-samples t-test was conducted. A two-sided Mann-Whitney U test was conducted for the PSQI scores due to the violation of normality. There was no significant difference between conditions for the MEQ or PSQI scores in any of the experiments.

**Supplementary Table 13.** MRS segments.

| Stage | Experiment 1 (n=19 subjects) | Experiment 3 (n=19 subjects) | Exp. 1 vs. Exp. 3 | NREM+REM vs. NREM-only |
|---|---|---|---|---|
| W | 2.3 ± 0.36 | 2.8 ± 0.38 | $U = 145.5$, $p = 0.296$ | Exp. 1: $U = 29.5$, $p = 0.193$ Exp. 3: $U = 29.5$, $p = 0.237$ |
| NREM | 6.7 ± 0.60 | 7.5 ± 0.65 | $U = 152$, $p = 0.411$ | Exp. 1: $U = 31$, $p = 0.262$ Exp. 3: $U = 33$, $p = 0.382$ |
| REM | 1.7 ± 0.34 | 1.9 ± 0.30 | $U = 33$, $p = 0.531$ | N/A |

A two-sided Mann-Whitney U test was used to compare the number of MRS segments between experiments as well as between the NREM+REM and NREM-only groups for each experiment because of the violation of normality shown by the Shapiro-Wilk test. The number of MRS segments for REM sleep includes only the data from the NREM+REM group because the number of REM sleep segments was not present for the NREM-only group. The results showed no significant difference in the number of MRS segments used for the experiments or groups. Note that data from the NREM+REM group only were used for REM segments, as the NREM-only group did not have REM sleep. Otherwise, data from both groups were combined.

**Supplementary Table 14.** Source data for Supplementary Table 1 - 13.