**Supplementary Methods**

*Participants*

Twenty-four right-handed (defined by the Edinburgh handedness questionnaire) participants (9 male, 20.7±0.5 years) were randomly and equally distributed between two experimental groups. An additional twenty-four right-handed participants (10 male, 21.2±0.6 years) were equally and randomly distributed between two control groups.

We also include overnight improvements (8pm to 8am, n = 10, 4 male, 18-24 yrs) and improvements which developed over 24-hr (8pm to 8pm, n = 8, 3 male, 18-24 yrs) drawn from an earlier study in which participants were trained, tested and following a night of sleep retested on a SRTT. The SRTT used in that earlier study and the current study are identical.

*Experimental Design*

Participants were trained and tested at 8am (Session 1), and subsequently retested at 8pm (Session 2). During the first session, participants practiced a sequence learning task (the SRTT, see below) which consisted of three blocks: a training block sandwiched between two test blocks (Figure 1). Following the final test block, participants were asked to either learn a list of words (Word list learning task, see below) or count the number of vowels within nonsense letter strings (Vowel counting task, see below).

After twelve hours, participants performed a single test block of the SRTT (second session, Figure 1). They were then asked, during a free recall test, to verbally describe the repeating sequence as an order of digits from 1 to 4. Each digit corresponded to one of the four screen positions used during the SRTT. Free recall was defined as the number of digits which were placed in the same order as they occurred in the sequence. For a digit to be scored as correct it had to appear within a segment of at least three correct consecutive digits (see, ). The difference between SRTT performance before (test, skill1) and after (retest, skill2) the interval provided a measure of off-line learning.
For one control group, the interval between session 1, containing the word list task, and session 2 was decreased to only 30 minutes. For the other control group, an interval of 4-hr was introduced between completing the final test block of session 1, and the word list task. Despite the introduction of this interval, session 2 was still completed 12-hr after session 1, at 8pm.

*The serial reaction time task (SRTT).*

We used a modified version of the SRTT. A solid circular visual cue (diameter 20mm, viewed from approximately 800mm) could appear at any one of four possible positions, designated 1 to 4, and arranged horizontally on a computer screen. Each of the four possible positions corresponded to one of the four buttons on a response pad (Cedrus, RB-410), upon which the participants’ fingers rested. When a target appeared, participants were instructed to respond as quickly and as accurately as possible by pressing the appropriate button on the pad. Having made the correct response, the cue on the screen disappeared and was replaced by the next cue after a delay of 400ms. If the participant made an incorrect response, the stimulus remained until the correct button was selected.

The position of the visual cues played out a repeating 12-item sequence (2-3-1-4-3-2-4-1-3-4-2-1). The training block had fifteen-repetitions (180 trials) of the twelve-item sequence. The two test blocks had only nine repetitions (108 trials). For all blocks, fifty random trials preceded and followed the sequential trials (Figure 1). Within these random trials there were no item repeats. Each set of random trials in the training and test blocks were unique. However, the random trials were identical across all groups.

Participants were instructed that a change in the color of the visual cues from black to blue marked the introduction of the repeating sequence. Earlier work has used a similar design to simultaneously engage declarative and procedural learning in the SRTT, and when this same task was used in an earlier study, off-line learning only occurred over a night of sleep, not over wake.
**Word list task**

A single word, from a list of 16 words (drawn from the California Verbal Learning Task), was presented on a computer screen for 2s. The word was then removed, and replaced by another word also drawn from the list of 16 words. This process continued until all 16 words had been presented. The same 16 words were presented individually and in the same order five times for each participant. At the end of each of these presentations, participants were asked to recall, in any order, as many of the words as possible. They were not prompted for particular words, nor were they told those words, if any, which they had failed to recall. Following the fifth recall, there was a ten-minute interval after which a free recall test of the word list was once again administered.

**Vowel counting task**

Participants were shown a nonsense string of letters, varying in length from 3-12 letters. Participants were asked to state the number of different vowels within a string. Each string was presented for 2s before being replaced by another nonsense letter string. This continued until 16 nonsense letter strings had been presented. This was a single iteration. In all there were a total of five iterations. Following the fifth iteration, there was a ten-minute interval, after which 16 nonsense letter strings were presented individually for 2s, and participants were asked to state the number of different vowels within each string. Each of the nonsense letter strings was unique.

**Data Analysis**

Response time was defined as the time taken to make a correct response. Any response time longer than 2.7 standard deviations (i.e. the top one percentile) from a participant’s mean was removed, as was any response time exceeding 2000ms. A learning score was calculated by subtracting the average response time of the final fifty sequential trials from the average response time of the final fifty random trials \(^4\). Skill before the interval \((skill_1)\) was calculated using the final test block of the first session, and skill after the interval \((skill_2)\) was calculated using the first and only test block of the second session (Figure 1). The difference \((skill_2 - skill_1)\) between these learning scores gave a measure of off-line learning. An ANOVA was used to compare initial skill and off-
line learning across the groups. Unpaired t-tests were used to compare between improvements in the experimental and control groups. To test for off-line learning within each group, paired t-tests were used to compare skill₁ against skill₂. Unpaired t-tests were used to make planned comparisons of sequence recall between groups. A linear regression was used to correlate off-line improvements with sequence recall. Finally, repeated measures ANOVA and paired t-tests compared between and within groups the number of incorrect responses made during the sequential trials at testing and retesting.

REFERENCES