Rule abstraction is facilitated by auditory cueing in REM sleep

REM sleep and abstract reasoning

Sofia Isabel Ribeiro Pereira[1#], Lorena Santamaria[1#], Ralph Andrews[1], Elena Schmidt[1], Mark Van Rossum[2], Penelope A. Lewis[1*]

[1]School of Psychology, Cardiff University Brain Research Imaging Centre, Cardiff, Wales, CF24 4HQ, UK

[2]School of Psychology and School of Mathematical Sciences, University of Nottingham, Nottingham, NG7 2RD, UK

#joint first authors

*Corresponding author LewisP8@cardiff.ac.uk

Abstract

Sleep facilitates abstraction, but the exact mechanisms underpinning this are unknown. Here, we aimed to determine whether triggering reactivation in sleep could facilitate this process. We paired abstraction problems with sounds, then replayed these during either slow wave sleep (SWS) or rapid eye movement (REM) sleep to trigger memory reactivation in 27 human participants (19 female). This revealed performance improvements on abstraction problems which were cued in REM, but not problems cued in SWS. Interestingly, the cue-related improvement was not significant until a follow up retest one week after the manipulation, suggesting that REM may initiate a sequence of plasticity events that requires more time to be implemented. Furthermore, memory-linked trigger sounds evoked distinct neural responses in REM, but not SWS. Overall, our findings suggest that targeted memory reactivation in REM can facilitate visual rule abstraction, although this effect takes time to unfold.

Keywords: sleep, rule abstraction, targeted memory reactivation, REM, SWS, synthetic visual reasoning task, event-related potentials, P300.

Significance Statement

The ability to abstract rules from a corpus of experiences is a building block of human reasoning. Sleep is known to facilitate rule abstraction, but it remains unclear whether we can manipulate this process actively and which stage of sleep is most important. Targeted Memory Reactivation (TMR) is a technique which employs re-exposure to learning-related sensory cues during sleep in order to enhance memory consolidation. Here, we show that TMR, when applied during REM sleep, can facilitate the complex recombining of information needed for rule abstraction. Furthermore, we show that this qualitative REM-related benefit emerges over the course of a week after learning, suggesting that memory integration may require a slower form of plasticity.

Introduction

Abstraction, or the process of formulating generalized ideas or concepts by extracting common qualities from specific examples, is a core component of fluid intelligence (Welling, 2007). Sleep has been suggested to play an active role in rule abstraction (for reviews see (Chatburn et al., 2014; Lerner and Gluck, 2019)). For instance, some experimental paradigms which probe rule abstraction such as statistical learning of tone transition patterns have been shown to benefit from slow wave sleep (SWS) (Durrant et al., 2011, 2013, 2016), whereas others, like the weather prediction task, seem to benefit from rapid eye movement sleep (REM)(Barsky et al., 2015). Rule-learning related neural patterns have even been shown to reactivate in the rat medial prefrontal cortex during SWS (Peyrache et al., 2009). However, the mechanisms supporting abstraction in sleep are unknown. It is unclear if one specific sleep stage is more important, and whether the benefit stems from memory reactivation or other types of processing in sleep.

Targeted memory reactivation (TMR) is a method for explicitly controlling memory reactivation in the sleeping brain (Oudiette and Paller, 2013). In TMR, sounds that have been simultaneously paired with recently learned material during wake are softly re-presented during subsequent sleep to trigger reactivation of the associated memories and boost consolidation. TMR is most commonly applied during non-REM (NREM) sleep, where it is known to strengthen memories (Rasch et al., 2007; Rudoy et al., 2009; Antony et al., 2012), but has also been linked to qualitative changes, such as the emergence of explicit knowledge of formerly implicit memories (Cousins et al., 2014). There is currently a debate in the literature regarding whether or not memories can be reactivated during REM sleep using TMR, with some studies reporting null findings (Rasch et al., 2007; Hu et al., 2020), and others reporting significant effects (Sterpenich et al., 2014; Hutchison et al., 2021; Picard-Deland et al., 2021). The present study aims to address this issue within the realm of rule abstraction, since the question of whether TMR can also boost this skill, in addition to memory consolidation, remains to be answered. It is also unclear whether rule abstraction would benefit most from

reactivation in SWS or in REM, given the proposed role of these sleep stages in memory restructuring (Landmann et al., 2015) and generalisation (Lewis and Durrant, 2011; Sterpenich et al., 2014; Pereira and Lewis, 2020). One study did apply SWS TMR to an abstraction task and suggest a benefit, but the lack of a non-cued control makes the results difficult to interpret (Batterink and Paller, 2017). Another study showed no effect of SWS TMR on generalisation (Witkowski et al., 2021), while in a third study, such stimulation appeared to produce a deficit in abstraction (Hennies et al., 2017). Nonetheless, SWS has been linked to positive effects in numerous abstraction-related tasks (see (Lerner and Gluck, 2019) for a review).

In the current report, we address the above questions by using TMR to reactivate rule abstraction problems in SWS and REM, with different problems cued in each stage. We used a visual abstraction task called the Synthetic Visual Reasoning Task (SVRT)(Fleuret et al., 2011) which requires participants to abstract rules that define 'families' of abstract visual patterns through trial and error exposure. For example, in the problem depicted in Figure 1, the rule is that each image contains two identical shapes. In training, participants are shown a series of images and asked to categorise them as belonging to the family in question or not. They are given feedback on each correct/incorrect categorisation. Each family of shapes is associated with a consistent reference image. At test, participants have to indicate whether or not a given sample image follows the same rule as the reference image for that particular problem. Because the impacts of TMR can last for up to a week(Hu et al., 2015), and may even amplify across this period(Groch et al., 2017), we re-tested our participants one week after the TMR manipulation.

Materials and Methods

Participants

Healthy young adults (mean age 22 years old, range = 19 – 30 years) were recruited online and through advertisements on the university campus to take part in this study. Participants filled out an online screening form and were excluded if: they had any diagnosed sleep, neurological or psychiatric disorders, were taking psychoactive medication, travelled more than two time zones or engaged in regular shift work in the two months prior to the experiment. Participants reported a regular sleep cycle over a four-week period prior to the experiment and were instructed to abstain from alcohol (24h) and caffeine (12h) prior to each visit to the laboratory, as well as daytime napping. Data from 27 individuals (19 females) were collected and used for behavioural analyses. One participant was excluded from the ERP analyses since, due to technical difficulties, no EEG triggers were recorded during TMR (n = 26). All participants signed informed consent and received monetary compensation for their participation. This study was approved by the ethics committee of the School of Psychology of Cardiff University.

Experimental design

The experiment was conducted according to a *within-subject* design (see Figure 1). Participants arrived in the evening (between 6 and 8pm) and were prepared for polysomnography recordings. Subsequently, participants performed a battery of pre-sleep cognitive testing. First, they performed the Image Familiarisation Task, where they passively saw all the images (either faces or landscapes) used in the SVRT. To ensure engagement, participants were instructed to press the space bar whenever a red dot appeared on the screen. After the Image Familiarisation Task, participants performed the Problem-Image Association Task, where they learned to associate each SVRT problem with a particular image of either a face or a landscape. These images were used to group the SVRT problems into 2 categories (category 1: problems paired with faces, category 2: problems paired with

landscapes). Next, participants performed the Synthetic Visual Reasoning Test (Fleuret et al., 2011), where they were required to categorize a series of samples from 16 problems as either in-class (following the rule) or out of class (not following the rule) (see Extended Data Figure 1-1). Each problem was always presented in combination with a specific image from one of the two possible categories (faces or landscapes) and with a 200ms sound. During training, participants learned through feedback and trial-and-error until they were able to correctly categorize the samples to 70% accuracy on each problem. During testing, they did not receive any feedback. The last task before sleep was the Problem-Sound Association Task, where participants were trained to recognize which sound had been paired with which problem, until they reached 100% accuracy. This task was introduced to guarantee that the effectiveness of TMR would not be compromised by a weak association between the sounds and their respective problems.

Next, participants went to sleep while non-obtrusive brown noise was continuously played throughout the night. For targeted memory reactivation, each category (sets 1 and 2 of problems paired with faces or problems paired with landscapes) was assigned to a sleep stage (either SWS or REM). Assignment of categories to the sleep stages was counterbalanced across participants. Within each category, half of the problems were cued during sleep and the other half served as a non-cued control (subsets A and B). Assignment of sets 1A, 1B, 2A and 2B to each sleep stage and cueing condition was counterbalanced across participants (see below which SVRT problems were included in each set). The sounds paired with problems assigned to the cued condition were played at the onset of either SWS or REM, as well as new, control sounds, not previously presented to the participant. Upon awakening (day 1), participants performed the Image Familiarisation task again, were wired down, showered, and then were retested on the SVRT. A week later (day 7) participants returned to the lab and were retested once again on the SVRT. Performance on the SVRT was assessed by the accuracy at each time point, and by the accuracy change (difference across time points).

148 All tasks were implemented in Matlab R2017b using Psychtoolbox 3 and displayed on a 1920

149 x 1080-pixel computer monitor.

Tasks

Image familiarisation task

152 This task consisted of 14 blocks of 8 trials (one per problem) for each one of the two categories

153 (i.e. 8 faces and 8 landscapes, for a total of 16 different images), amounting to 112 image

154 presentations per category (224 in total). A variable inter-trial interval was set between 1 and

155 2 seconds. Participants were asked to press the space bar whenever a red dot appeared on

156 the screen. The red dot was set to appear randomly once every 8 trials. The task was

157 administered in the evening and again in the morning.

Problem-Image Association task

159 This task was designed to help participants learn to associate each SVRT problem and its

160 corresponding sound with a particular image (either a face or a landscape). It consists of 2

161 phases: learning and test. For each participant, the images and sounds were randomly

162 assigned to the SVRT problems. During learning, participants performed 3 blocks of 16 trials

163 (one per problem) where they passively viewed the reference representation of any given

164 SVRT problem on the left-hand side of the screen and the image it was paired with (either a

165 face or a landscape) in the centre of the screen, while the 2 second sound paired with that

166 problem-image dyad was played. Participants were instructed to press the space bar if a red

167 dot appeared on the screen. The red dot appeared randomly once per block. In the test phase,

168 participants saw the reference representation in the centre of the screen and heard the same

169 sound that had been paired with it during learning, but now trimmed to only 200 ms. Next, two

170 images appeared on the screen and the participant had to indicate which one had been paired

171 with that particular problem-sound dyad. The test was repeated until participants reached 75%

172 accuracy.

173   These two tasks, image familiarisation and problem-image association, were added to the

174   experimental design in order to facilitate use of machine learning classification algorithms to

175   detect replay. We performed extra checks to certify that image category was not influencing

176   the SVRT task, see results.

177   *Synthetic Visual Reasoning Test (SVRT)*

178   The SVRT task requires participants to indicate whether or not a given sample image follows

179   the same rule as the reference image for that particular problem (both sample and reference

180   images were displayed simultaneously). The rule governing each problem had to be

181   discovered through trial and error during training. We measured accuracy as the ability to

182   correctly categorize sample images according to whether they followed, or broke, the rule for

183   that problem (Figure 1). Feedback was given after each trial, informing participants whether

184   or not their categorisation of the sample image was correct. For more examples of sample

185   images and rules, please refer to Extended Data Figure 1-1. Each problem was presented in

186   conjunction with a picture of a face or a landscape, to boost the chances of eliciting classifiable

187   EEG patterns, as has been done for objects and scenes (Cairney et al., 2018), and for animals,

188   tools, faces and buildings (Shanahan et al., 2018). Participants were trained on 16

189   categorization problems, half of which were subsequently used to test the impact of TMR in

190   SWS (4 were cued in SWS and 4 were used as a control), and the other half (4 cued and 4

191   control) were used to test the impact of TMR in REM.

192   The test phase consisted of 5 trials for each problem. Out of a pool of 200 images per problem,

193   (100 following the rule and 100 not following the rule), 5 images were randomly selected for

194   each test (pre-sleep, Day 1 and Day 7).

195   During both training and test phases, a time limit for each response was set to 6 seconds,

196   after which the next trial would start. After each block (i.e. problem) there was a 15 second

197   rest break. The order of problem presentation was randomized for each participant. Each trial

198   began with the presentation of that problem's reference representation on the left-hand side

199 of the screen, the image it had been paired with (either a face or a landscape) in the centre

200 for the screen, and the 200ms sound that these images were associated with. Then, the image

201 to be categorized was displayed on the right-hand side of the screen. Participants were

202 required to press 1 if the image to be categorized was in class (satisfied the rule) or to press

203 9 if it was out of class (did not satisfy the rule). Performance on the SVRT was assessed by

204 the change in accuracy overnight (post-sleep day 1 – pre-sleep), across the week (post-sleep

205 day 7 – post-sleep day 1). Performance was not affected by the category of the image paired

206 with each problem (i.e. face or landscape(all t-tests p > 0.4, uncorrected)).

207 *Problem-Sound Association task*

208 This task was designed to ensure that participants were able to correctly identify all sound-

209 problem dyads introduced while performing the SVRT before sleep, which could otherwise

210 compromise the effectiveness of TMR. Again, the reference representation was presented in

211 combination with its corresponding face or landscape image. Next, two 200ms long sounds

212 were played and the participant indicated which one had been paired with that problem-image

213 dyad. The test was repeated until participants reached 100 % accuracy.

214 **Stimuli**

215 All sounds were obtained from an online repository (www.freesound.org). Initial sounds (2

216 seconds long; learning phase of the Sound-Problem Association Task) were trimmed into

217 200ms long sounds using the software Audacity. A pool of sounds was used for each category

218 (faces/landscapes), from which sounds were randomly selected and assigned to a specific

219 SVRT problem. For faces, generic object sounds were used and for landscapes, generic

220 nature sounds were used, such as a bird chirping or the wind blowing. For each category

221 (faces or landscapes) a group of 12 similar but easily distinguishable sounds was selected

222 and from this pool, 8 sounds were randomly paired with an image and used in the SVRT task

223 while the remaining 4 sounds were used as controls during TMR. Sounds for faces and

landscapes were matched in duration, and all were played at the same volume within each participant.

The images of faces were obtained from the Karolinska Directed Emotional Faces (KDEF) (Lundqvist et al., 1998). Only faces of females with a neutral facial expression at a straight angle were chosen. The images of landscapes were obtained from an online repository (www.freeimages.com). All images were edited into grayscale and resized (faces: 325 x 435 pixels; landscapes: 435 x 325 pixels) using the software GIMP.

TMR protocol

Audio cues were embedded in brown noise in order to decrease the likelihood that the TMR sounds would elicit an arousal. Brown noise was played throughout the entire night while the cues were only presented when SWS or REM was identified online by the experimenter. Both stimuli (audio cues and brown noise) were played through loud speakers placed behind the participant's bed. The sound volume was manually adjusted for each participant before sleep according to their comfort level. Each cue (either experimental, e.g. paired with a learned rule or control, with no rule associated) was played twice in a row before the next cue was played. All cues were played 4 seconds apart from each other. One loop of cueing consisted of all 8 cues (4 control and 4 experimental) played twice (16 sound presentations). The order of cue presentation was randomized at each iteration of the loop. A total of 14 loops was played in each sleep stage (corresponding to approximately 15 min of cueing), adding up to 28 repetitions of each individual sound and 112 cueing events in each condition (control or experimental). Even though SWS usually occupies a larger proportion of the night than REM (and would thus allow for an extended cueing time), we wanted to ensure that we would be able to deliver the same amount of cueing in both sleep stages, and therefore we opted for limiting cueing to ~ 15 min. Cueing was initiated in the first episode of SWS and REM and was interrupted whenever an arousal or sleep stage transition was identified. In one participant, only 7 out of the 14 loops of REM cueing were completed, due to short sleep duration (n = 1) and in another participant only 8 out of the 14 loops of SWS cueing were completed, due to

251   light sleep throughout the night (n = 1). These participants were not excluded from any

252   analyses. Note that cueing varied between participants, depending on whether or not they

253   obtained ~15 min of uninterrupted SWS and REM, such that for some cueing was finished

254   within the first NREM-REM cycle while for others additional cycles were needed. No significant

255   correlations were found between number of cues delivered in SWS or REM and subsequent

256   performance (all $p > 0.1$). Following offline sleep scoring, cueing accuracy (calculated as the

257   percentage of cues delivered in the intended sleep stage) was determined: 94.44 % for SWS

258   and 93.72 % for REM. Regarding continuity (i.e. whether or not TMR was completed within on

259   sleep cycle, SWS TMR was continuous for 19 participants out of 26 participants and REM

260   TMR was continuous for 1 out of 26 participants only. This is to be expected, since we initiated

261   REM TMR at the onset of the first REM episode, which tends to be very short and our entire

262   cueing procedure required at least 15 min to complete, if uninterrupted. Given this distribution

263   of the data, it is not possible to estimate if the TMR effect differed depending on whether

264   cueing was continuous or discontinuous.

265   **EEG recordings and sleep analysis**

266   EEG was recorded using BrainVision software during the Image Familiarisation task (in the

267   pre-sleep evening and morning of post sleep day 1) and during sleep. Recordings were made

268   at 500 Hz from 22 scalp locations on the standard 10/20 layout (Fz, F3, F4, FC1, FC2, FC5,

269   FC6, Cz, C3, C4, CP5, CP6, Pz, P3, P4, P7, P8, PO3, PO4, Oz, O1 and O2), referenced to

270   the mastoids. Impedances were kept below 5 kΩ. Electrooculogram (EOG) and

271   electromyogram (EMG) signals were also recorded from electrodes next to each eye and 2

272   electrodes on the chin, respectively. Sleep scoring was accomplished using the guidelines

273   from the American Association of Sleep Medicine (AASM, v. 2.5), within a custom-made script

274   implemented in Matlab. Offline scoring was performed by two independent raters, blind to

275   when cueing occurred, achieving an 88% agreement rate. Discrepancies were resolved by

276   one of the raters.

277   Spindles and slow oscillations were detected from all channels using the SpiSOP toolbox

278   version 2.3.8.3 (available at https://www.spisop.org/), with the spindle detection algorithm

279   based on (Molle et al., 2002). Centre frequencies of fast and slow spindles were visually

280   determined for each participant and used to define the finite impulse response (FIR) filter

281   (center frequency 13.29Hz (std: 0.69)). The root mean square (RMS) of the filtered signal was

282   computed using a 0.2s time window and smoothed by a moving average of another 0.2s

283   window. Any event that surpassed the 1.5 SD of the RMS signal was considered a candidate

284   spindle. To fit the spindle detection criteria, the candidate events had to last between 0.5s and

285   3s. Because we had no *apriori* hypothesis about specific channels, all correlations were made

286   with the average across channels.

287   Similarly slow oscillation detection is based on (Mölle et al. 2002) but also see (Ngo et al.

288   2013). Prior to the actual detection, the signal is high pass filtered (IIR by default) then low

289   pass filtered (FIR) to contain frequency components observed in slow oscillations in a

290   specified band (0.3 to 3.5 Hz). Then all the time intervals with consecutive positive-to-negative

291   zero crossings are marked. Only intervals with durations corresponding to a minimum (set to

292   0.5Hz) and maximum (set to 1.11Hz) slow oscillation frequency are considered as putative

293   slow oscillations. The threshold for negative peaks is set to 1.25 and for negative to positive

294   peaks amplitude was also set to 1.25 (default parameters).

295

296   **EEG pre-processing**

297   First, the data was high-pass filtered at 0.3 Hz and low-pass filtered at 35 Hz. Then, the

298   continuous EEG was epoched into trials from 1 s before to 3 s after sound cue onset (since

299   the cues were 4 s apart). Noisy channels were repaired by interpolating data from

300   neighbouring electrodes and trials containing arousals or movement artefacts (as determined

301   during sleep scoring) were removed. Finally, any remaining noisy trials were manually

removed following visual inspection. The number of trials included in the final analysis for each participant, sleep stage and condition are presented in Extended Data Figure 2-3.

Baseline correction was performed on the single trial level using the entire trial length [-1 3] (Grandchamp and Delorme, 2011). Trials were then separated into conditions (control and experimental) and sleep stages (SWS and REM). One participant was excluded from all analyses, since they did not have EEG triggers during TMR (final n = 26).

EEG analysis

Event-related potentials (ERPs) analyses were carried out in Fieldtrip (Oostenveld et al., 2011) (available at: http://www.fieldtriptoolbox.org/). ERPs were calculated for each condition and sleep stage, and compared within subjects and between conditions, across all channels, within a time window from 0 to 2000ms (not averaged).

ERPs of control and experimental sounds were compared using Monte-Carlo cluster permutation tests, corrected for multiple comparisons (Maris and Oostenveld, 2007). The cluster alpha was set to 0.05 and 150000 randomizations were carried out for every test. Clusters were considered significant at $p < 0.025$ (two tailed). Similar parameters were set-up for time-frequency analysis for each frequency band of interest: theta (4 to 8Hz), spindles (9 to 15Hz) and low-beta (12.5 to 16Hz). More specifically, the time-frequency cluster permutation analysis was calculated using the average across trials for each participant in the window of interest (0 to 2s). The statistical analysis was performed for experimental vs control sounds in SWS, REM and also for their interaction (SWS difference vs REM difference, where difference was calculated as experimental minus control sounds) for each frequency band. The minimum number of channels to form a cluster was set to 2, the number of randomisations set to 250000 and the cluster alpha at $p=0.025$ (two-tailed).

To determine whether stimulation lead to a change in spindles or slow oscillations, we calculated the number and duration of spindles and slow oscillations per condition (experimental and control sounds). We then compared these between conditions using a

328   cluster permutation analysis. The cluster alpha was set to 0.05 and 250000 randomizations

329   were carried out for every test. Clusters were considered significant at p < 0.025 (two tailed).

330   Finally, we sought to detect memory reactivation after our TMR cues using an EEG classifier.

331   Thus, ERP values were used as features to feed a linear Support Vector Machine (SVM). To

332   avoid overfitting, we used 5-fold validation repeated twice. As a performance metric we used

333   the traditional accuracy but also area under the curve. The classification was performed

334   separately for SWS and REM stages for each participant. Statistics were performed at a group

335   level to check if for any above-chance time-cluster. No significant cluster was found for either

336   of the performance metrics or for either sleep stage.

337   **Statistical analyses**

338   Performance change on the SVRT was compared using a repeated measures ANOVA with

339   between-subjects factors sleep stage (SWS/REM), cueing condition (cued/non-cued) and

340   session (overnight/across the week) as repeated factor. We ran an outlier analysis using the

341   ROUT method (Q = 1%) and identified two outliers on the SWS cued group. Upon removal of

342   these outliers, the results remained the same as those in Figure 2A, where no significant

343   differences were found between overall performance change on SWS cued and non-cued

344   problems (t (1,24) = 1.132, p = 0.269).

345   Descriptive statistics (mean, standard deviation, standard error of the mean and confidence

346   intervals) are presented in Figure Extended Data Figure 2-4 The combined performance

347   change was compared between non-cued and cued conditions using paired t-tests. Pearson's

348   correlations were calculated between the combined performance change and the average

349   number of slow oscillations and spindles in frontal, central and parietal derivations. Data are

350   presented as mean ± SEM and we report eta squared (η^2) and Cohen's *d* as effect size

351   estimates for significant findings.

352   Statistical analyses of the behavioural data were conducted on JASP 0.10.2.0 while statistical

353   analyses of EEG data were conducted on Matlab R2017b using the Fieldtrip toolbox (version

354   20190904).

355

Results

TMR in REM improves rule abstraction

We examined baseline performance (pre-sleep) using an ANOVA with the factors cueing condition (cued/non-cued) and Sleep stage (SWS/REM). No differences or interaction were found (smallest p=0.666). Refer to Figure 2.B and Extended Data Figure 2-1a for full statistical details.

To assess the impact of cueing, upon consolidation across a retention interval, we compared SVRT performance change (overnight accuracy change: post-sleep day 1 - pre-sleep; and across a week: post-sleep day 7 - post-sleep day 1) using a repeated measures ANOVA with factors sleep stage (SWS and REM), cueing condition (cued and non-cued), and retention interval (overnight and across a week post-sleep) as repeated measure. This showed a significant sleep stage*cueing condition interaction ($F_{(1,26)}$ = 6.091, p = 0.020, η^2 = 0.013), with no other factor or interaction being significant (smallest p=0.128, Figure 2A, Extended Data Figure 2-1b). This indicates that cueing had different effects when applied in SWS and REM. To investigate this, we conducted a simple main effects test (sleep stages x cueing), which revealed better performance in the cued condition for REM than SWS ($F_{(1,26)}$ = 4.463, p = 0.044), with no differences between SWS and REM in the non-cued control condition ($F_{(1,26)}$ = 0.774, p = 0.387; Figure 2A). This result could suggest that cueing benefited rule abstraction when delivered during REM sleep, but not SWS.

To better understand this pattern of results, and also to gain statistical power, we next analysed each sleep stage separately using a 2-way ANOVA with factors cueing condition (cued and non-cued) and retention interval (overnight and across a week post-sleep). For SVRT problems cued in SWS, there was no effect of cueing, session or interaction between these (smallest p=0.198). For problems cued in REM sleep however, we found a significant cueing effect ($F_{(1,26)}$ = 7.930, p = 0.009, η^2 = 0.019), indicating that performance improvements were superior for cued problems, compared to non-cued problems. There was no effect of

382   session or cueing*session interaction (*smallest p=0.231*). To further understand the origin of

383   the cueing effect in REM sleep we performed a paired t-test (cued vs non-cued) on accuracy

384   at each session (Pre-sleep, post-sleep day1 and post-sleep day 7), Figure 2B and Extended

385   Data Figure 2-3 for full statistical results. Accuracy was superior for REM cued problems, as

386   compared to non-cued ($t_{(26)}$ = 3.357, p = 0.002, Cohen's d = 0.646) only at Post-sleep day 7.

387   Overall, these findings suggest that reactivating problems during REM leads to a significant

388   advantage in rule knowledge after seven days and nights.

389

390   **Event-related potentials in REM differ between control and experimental sounds**

391   To examine neural processing associated with TMR cues, we plotted sound-evoked ERPs for

392   each sleep stage of cueing (SWS and REM) and sound category (control and experimental)

393   at Cz for illustration purposes, see Figure 3. Topographies showing the spatial distribution of

394   significant channels over time are available in the (Figure 4 for all EEG channels). We

395   analysed a large time window (0–2000ms), which includes all known auditory event-related

396   potentials (Winkler et al., n.d.) and has previously been associated with processing auditory

397   stimuli in both NREM and REM sleep (Campbell and Muller-Gass, 2011). To determine

398   whether the response to control and experimental sounds differed in each sleep stage, we

399   performed a cluster analysis on the ERP amplitudes (all channels, not averaged). This

400   revealed a significant difference between experimental (familiar) and control (new) sounds in

401   REM sleep (cluster corrected for multiple comparisons, p=0.048), but not in the SWS (all p >

402   0.05). This negative cluster ranges from 228ms to 400ms. The elicitation of a larger ERP

403   amplitude for new sounds than for familiar sounds demonstrates an ability to detect novelty.

404   Our observation of this response in REM but not SWS is in keeping with prior literature

405   showing greater responsivity in REM compared to SWS (see(Ibáñez et al., 2009) for a review).

406   To probe the data further, we performed a time-frequency analysis per sleep stage in the same

407   time window (0-2000ms) choosing relevant frequency bands based on previous work on SWS:

408 theta-band (4-8Hz) and spindle band (9-15Hz), and lower beta band (13-16Hz) for REM sleep.

409 Cluster statistics revealed nothing significant for either frequency band or sleep stage

410 (smallest p-value 0.052). Full list of results in Extended data Figure 4-1.

411

412 Does cueing in each sleep stage interfere with consolidation of cueing in the other?
413
414 Because we applied TMR in both SWS and REM (though stimulating different problems in

415 each stage) we were interested to know whether TMR in REM might have obscured or

416 interfered with the effects of TMR in SWS. In the case of direct interference, we might expect

417 a negative correlation between the extent to which participants benefit from REM TMR and

418 the extent to which they benefit from SWS TMR. To test for this, we looked for a relationship

419 between performance on problems cued in SWS and REM in two different ways, using

420 overnight gain and using TMR cueing benefit. Thus, we ran a correlation between overnight

421 performance change (difference between post-sleep and pre-sleep) for problems cued in SWS

422 and overnight performance change for problems cued in REM. This showed no correlation (r

423 = -0.162, p = 0.420). Next, we calculated the cueing benefit (difference between performance

424 on cued and non-cued problems) for SWS-related problems and REM-related problems at

425 each session and across sessions, to check if TMR-related improvements in REM problems

426 were obtained at the expense of cueing benefit in problems cued in SWS. This showed no

427 significant relationships (p > 0.05, uncorrected; Table 2). These results show that the extent

428 of TMR related consolidation in REM doesn't predict any specific deficit in the benefit accrued

429 from equivalent cues in SWS.

430

431 There is no relationship between time spent in non-manipulated REM sleep and

432 performance on problems cued in SWS

433 It could be argued that successive TMR in SWS and REM might have curtailed the amount of

434 non-manipulated REM available to further advance any consolidation processes initiated by

435 TMR in SWS, thus disrupting any potential benefits from this manipulation. We inspected sleep

436   architecture in relation to TMR and found that 25 out of 26 participants had a period of non-

437   manipulated REM sleep after REM cueing had terminated: an average of 65.9 min (ranging

438   from 24 min to 117.5 min). Furthermore, the amount of non-manipulated REM sleep in each

439   participant was not correlated with performance on SWS cued problems on either post-sleep

440   day 1 (r = 0.284, p = 0.160) or post-sleep day 7 (r = 0.166, p = 0.419).

441

Relation between rule abstraction and NREM graphoelements

443   Sleep architecture data from all 27 participants is presented in Table 1.

444   Slow oscillations and sleep spindles are thought to mediate TMR-related benefits to memory

445   consolidation(Schouten et al., 2017; Cairney et al., 2018; Göldi et al., 2019). In order to

446   determine if the same was true for rule abstraction, we counted the number of slow oscillations

447   and sleep spindles in NREM sleep for each participant and checked for correlations between

448   each of these and the SVRT performance change for problems cued in SWS and REM, as

449   well as the control non-cued problems for each sleep stage. In line with the observation that

450   TMR in SWS did not improve rule abstraction, we found no correlation between performance

451   on the SVRT task and either spindles or slow oscillations (all p >= 0.1, uncorrected, Table 4).

452   Next, we wanted to determine whether TMR cueing altered spindles or slow oscillations in a

453   way that related to subsequent changes in performance on our task. We thus calculated the

454   number and duration (samples) of spindles and slow oscillation in the 3 second epoch

455   following TMR stimulation for each condition (experimental and control). No significant results

456   were found for spindles (smallest p value=0.06, see topography in Figure 5). But two

457   significant clusters were found for the number of SOs. One in the left hemisphere, t=-9.08 p

458   value=0.007, and one on the right hemisphere (t=-6.50, p=0.012), see Figure 5. Both indicated

459   a higher number of SO after control than experimental sounds. We then correlated the mean

460   number of SOs detected in each cluster with behavioural performance change for items (cued

461   in REM/SWS and non-cued for both stages) both overnight and over the subsequent week

462   and for both cued and non-cued items. This revealed a significant positive relationship

463   between both the right hemispheric cluster (Rho = .44, p=0.03) and the left hemispheric cluster

464   (Rho = .42, p= 0.04), uncorrected. Overall, these data appear to suggest that cueing with the

465   experimental TMR tone lead to a reduction in SOs over these electrodes and this seems to

466   be associated with TMR benefit, although the correlations do not survive correction for multiple

467   comparisons. However, because we had no *apriori* hypothesis to this effect, and the

468   correlations do not survive correction for multiple comparisons, we feel this should be treated

469   with caution.

470

471   **Image category did not affect SVRT performance**
472
473   To determine whether being associated with the face/object sounds versus the

474   landscape/nature sounds had any impact on behaviour, we directly compared performance

475   on problems associated with faces and landscapes, irrespective of sleep stage or cueing

476   condition. There were no differences in performance between the two. We conducted a two-

477   way repeated measures ANOVA on the raw accuracy values with the factors category: (faces

478   and landscapes) and session: (pre-sleep, post-sleep day 1 and post-sleep day 7). There was

479   no effect of category ($F_{(1,26)}$ = 0.362; p = 0.553; η^2 = 0.003) or session ($F_{(1,26)}$ = 2.054; p =

480   0.139; η^2 = 0.007) , and no interaction ($F_{(1,26)}$ = 0.253 ; p = 0.778; η^2 = 0.001). The same

481   analysis was conducted on the performance changes (overnight, over a week and overall

482   change), with Greenhouse-Geisser sphericity correction. Similarly, no effect of category ($F_{(1,26)}$

483   = 0.365; p = 0.551; η^2 = 0.004) or session ($F_{(1,26)}$ = 0.610; p = 0.480; η^2 = 0.004) was found,

484   and there was no interaction ($F_{(1,26)}$ = 0.165; p = 0.729; η^2 = 0.002). We ran paired t-tests

485   between the same time points in each category (e.g. Faces at pre-sleep vs Landscapes at

486   pre-sleep). No differences were found (all p > 0.4, uncorrected).

487

488   **Discussion**

489   This study shows that rule abstraction, one of the building blocks of human reasoning, can be

490   facilitated by applying targeted memory reactivation during sleep. Interestingly, when different

problems were cued in SWS and REM within the same night, the problems cued in REM benefitted from offline rehearsal, shedding light on a possible role for previously detected reactivation during REM (Maquet et al., 2000; Louie and Wilson, 2001; Mainieri et al., 2019). Furthermore, we found that REM TMR mediated facilitation of abstraction requires time to emerge, since cued problems have a significant advantage over non-cued problems one week after the manipulation. This is important, because it joins a small but growing literature suggesting that some sleep-related memory benefits may require more than just one episode of sleep to emerge (Groch et al., 2017; Cairney et al., 2018).

Abstraction underpins the ability to categorise items and generalize rules to new, never before seen exemplars. This is a core component of fluid intelligence(Otero, 2017), and is particularly important when one is faced with a new problem that cannot be solved exclusively by prior knowledge. Our data appear to show a dissociation between REM and SWS, with TMR in the former but not the latter facilitating performance on a complex task requiring rule abstraction and pattern categorization. Un-manipulated SWS has been shown to be involved in both quantitative (Rasch and Born, 2013) and qualitative changes to recently encoded memories (Wagner et al., 2004; Lau et al., 2010; Durrant et al., 2011, 2013; Wilhelm et al., 2013; Kirov et al., 2015), while REM has been suggested to be more involved with qualitative changes, such as forming unexpected links between different memories or concepts (Lewis et al., 2018). This possibility is supported by studies showing that REM duration predicts visual abstraction (Lutz et al., 2017), category learning (Djonlagic et al., 2009), lexical integration (Tamminen et al., 2017) and grammar learning (Batterink and Paller, 2017), all of which are highly integrative forms of memory. Our finding with respect to REM is also in line with a recent review suggesting that abstraction of explicit rules based on prior knowledge is often linked to REM sleep (Lerner and Gluck, 2019), and extends these ideas by providing clues to the underlying mechanisms of REM-dependent rule abstraction. In addition, one study demonstrated that TMR in SWS can actually impair the abstraction of grammar-like transition statistics(Hennies et al., 2017), suggesting that promotion of memory for specific episodes through reactivation

518 in SWS may disrupt the abstraction of generalised statistics. Taken together with this literature,

519 our findings suggest that REM TMR may have the capacity to directly promote abstraction.

520 Supporting this, studies using REM TMR to investigate qualitative changes, such as the

521 affective tone of emotional memories (Rihm and Rasch, 2015; Lehmann et al., 2016) and the

522 generalization/integration of pictures with emotional content (Sterpenich et al., 2014), typically

523 do find a benefit from REM TMR, as did our current study. If abstraction-like processing turns

524 out to be the main function of REM for memory, that could explain why most REM TMR studies

525 have shown little or no benefit to memory consolidation (for a meta-analysis see (Hu et al.,

526 2019)), since such studies typically assessed quantitative, rather than qualitative changes,

527 and thus do not test abstraction.

528 In the current study, while TMR in REM facilitated rule abstraction, TMR in SWS did not. Given

529 this result, it might be tempting to conclude that TMR in SWS does not facilitate this kind of

530 abstraction. However, we cannot exclude the possibility that cueing problems in SWS

531 triggered a consolidation process which would have facilitated abstraction, but which was

532 disrupted by subsequent cueing in REM. We ran several analyses to investigate this

533 possibility and found that there is no relationship between the extent to which SVRT

534 performance benefitted from cueing in REM and cueing in SWS. We also found that the vast

535 majority of participants had epochs of non-manipulated REM sleep after REM cueing had

536 ceased, which presumably provided an opportunity for items that had been cued in SWS to

537 continue their consolidation in REM as needed. Nonetheless, we still cannot rule out some

538 kind of interference and thus remain cautious in our interpretation. We therefore conclude

539 only that REM TMR is sufficient to start a consolidation process which facilitates rule

540 abstraction and cannot draw conclusions about the impacts of SWS TMR on this process

541 based on the current data alone.

542 Regarding the timing of the TMR effects, our data suggest that the impact of TMR may

543 continue to unfold for at least a week, with performance on cued and non-cued problems only

544 becoming significantly different after that temporal delay. Notably, we did not test performance

545   between days one and seven, so we do not know how quickly this process unfolds. If

546   qualitative changes in memory representations, such as abstraction, require longer periods of

547   time to evolve (Sterpenich et al., 2014; Lutz et al., 2017), then they may escape detection by

548   the commonly used 12 hour test-retest paradigm. Prior studies have considered longer test

549   periods and have shown that TMR-related benefits sometimes disappear over a week

550   (Shanahan et al., 2018), but can also persist over this period (Hu et al., 2015; Groch et al.,

551   2017; Simon et al., 2018). Our current study builds on these reports by showing that the benefit

552   to abstraction which was not significant at day one post-sleep became significant by day

553   seven. This is in keeping with a study of emotional processing, which showed that the impact

554   of NREM TMR on emotional content was amplified across a week (Groch et al., 2017), and

555   also with our own work on the serial reaction time task which shows that benefit from TMR

556   can emerge after 10 days or more (Rakowska et al., 2021).

557   Building on a model of synaptic plasticity across brain states (Redondo and Morris, 2011; Seibt

558   and Frank, 2019), we have recently proposed a series of plasticity-related events that take

559   place in both NREM and REM which could explain why the effect of sleep on memory

560   consolidation may require extended periods of time before it becomes detectable (Pereira and

561   Lewis, 2020). According to a recent framework (Seibt and Frank, 2019), neuronal ensembles

562   associated with the task are tagged during wakeful encoding. During subsequent NREM

563   reactivation, mRNAs or other Plasticity-Related Products (PRPs) are captured by these

564   tagged synapses. Finally, in subsequent REM, these PRPs are translated into proteins which

565   enable synapses to undergo intense remodelling. In light of our current results, we speculate

566   that applying TMR in REM might potentially bypass the need for PRP capture in NREM,

567   instead promoting PRP capture and translation at task-related synapses. Given the time-

568   consuming nature of these processes, multiple nights of sleep could be required before

569   measurable behavioural effects emerge. Of course, this does not explain why TMR cueing in

570   SWS, which might reasonably be expected to result in extra PRP capture by task-related

571   synapses, did not result in a behavioural benefit. We can only speculate that such PRP

572 capture is not sufficient in the case of our abstraction task. Alternatively, it is also possible

573 that cueing in REM subsequent to SWS somehow interfered with consolidation such that

574 PRPs capture during SWS cueing were not subsequently translated. More work will be needed

575 to disentangle such effects.

576 Our ERP analysis complements our behavioural findings by revealing differential neural

577 responses to experimental and control stimuli in REM, but not SWS. These differential

578 responses were found between 228 to 400ms post cue onset, a time window during which

579 auditory stimuli are known to be extensively processed in both NREM and REM sleep

580 (Campbell and Muller-Gass, 2011) and which is also associated with the P300 component

581 (Picton, 1992). The P300 is thought to reflect higher order cognitive processing related to

582 selective attention and resource allocation, with its amplitude proportional to the amount of

583 attentional resource recruited for scrutiny of a given stimulus (Ibáñez et al., 2009). The P300

584 has also been detected during REM, with larger peak amplitudes occurring for rare sounds in

585 the oddball paradigm (Cote and Campbell, 1999). Our data mirror this result by showing that

586 'new' control sounds elicited greater P300 waves than 'familiar' task-related sounds.

587 Interestingly, the P300 has been found in response to hearing one's own name in REM sleep,

588 but not in response to hearing another name. This could indicate that some level of cognitive

589 processing persists during REM (Bastuji et al., 2002). The fact that we observed a difference

590 between familiar and unfamiliar P300 responses in REM but not in SWS, is therefore in

591 keeping with the literature. Other authors have interpreted such results as suggesting that

592 stimuli are processed at a deeper, more cognitive, level during REM (see (Ibáñez et al., 2009)

593 for a review).

594

595 **Conclusion**

596 In sum, we found that TMR in REM is sufficient to benefit a visual reasoning task commonly

597 used in the field of Artificial Intelligence (Fleuret et al., 2011; Ellis et al., 2015), but never before

tested in a sleep study. Furthermore, ERPs suggested a deeper level of processing in REM than SWS, and behavioural findings suggest that the process started by TMR in REM requires more than one night of sleep to unfold. These findings open exciting new avenues for exploring TMR as a tool to enhance higher order cognitive functions such as abstraction, a core component of fluid intelligence and creativity.

Acknowledgments

Author Contributions

SIRP, PL and MVR designed the experiments, SIRP, RA and ES collected the data, SIRP and LS analysed the data and all authors wrote the manuscript.

Declaration of Interests

The authors declare no competing interests.

Source data

620 The full dataset presented here, including demographics, behavioural and EEG data, as well

621 as the Matlab scripts used in the ERP analyses, is available at 10.5281/zenodo.7215812.

622

623 **References**

624 Antony JW, Gobel EW, O'Hare JK, Reber PJ, Paller KA (2012) Cued memory reactivation
625   during sleep influences skill learning. NatNeurosci 15:1114–1116.

626 Barsky MM, Tucker MA, Stickgold R (2015) REM sleep enhancement of probabilistic
627   classification learning is sensitive to subsequent interference. Neurobiol Learn Mem
628   122:63–68.

629 Bastuji H, Perrin F, Garcia-Larrea L (2002) Semantic analysis of auditory input during sleep:
630   Studies with event related potentials. Int J Psychophysiol 46:243–255.

631 Batterink LJ, Paller KA (2017) Sleep-based memory processing facilitates grammatical
632   generalization: Evidence from targeted memory reactivation. Brain Lang 167:83–93.

633 Cairney SA, Guttesen A á. VV, El Marj N, Staresina BP (2018) Memory Consolidation Is
634   Linked to Spindle-Mediated Information Processing during Sleep. Curr Biol 28:948-
635   954.e4 Available at: http://linkinghub.elsevier.com/retrieve/pii/S0960982218301532.

636 Campbell K, Muller-Gass A (2011) The extent of processing of near-hearing threshold stimuli
637   during natural sleep. Sleep 34:1243–1249.

638 Chatburn A, Lushington K, Kohler MJ (2014) Complex associative memory processing and
639   sleep: A systematic review and meta-analysis of behavioural evidence and underlying
640   EEG mechanisms. Neurosci Biobehav Rev 47:646–655.

641 Cote KA, Campbell KB (1999) P300 to high intensity stimuli during REM sleep. Clin
642   Neurophysiol 110:1345–1350.

643 Cousins JN, El-Deredy W, Parkes LM, Hennies N, Lewis P a. (2014) Cued Memory
644   Reactivation during Slow-Wave Sleep Promotes Explicit Knowledge of a Motor
645   Sequence. J Neurosci 34:15870–15876 Available at:
646   http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.1011-14.2014 [Accessed
647   November 27, 2014].

648 Djonlagic I, Rosenfeld A, Shohamy D, Myers C, Gluck M, Stickgold R (2009) Sleep
649   enhances category learning. Learn Mem 16:751–755.

650 Durrant SJ, Cairney SA, Lewis PA (2013) Overnight Consolidation Aids the Transfer of
651   Statistical Knowledge from the Medial Temporal Lobe to the Striatum. Cereb Cortex
652   23:2467–2478.

653 Durrant SJ, Taylor C, Cairney S, Lewis PA (2011) Sleep-dependent consolidation of
654   statistical learning. Neuropsychologia 49:1322–1331.

655 Durrant SJSJ, Cairney SASA, Lewis PAPA (2016) ScienceDirect Cross-modal transfer of
656   statistical information benefits from sleep. CORTEX 78:85–99 Available at:
657   http://dx.doi.org/10.1016/j.cortex.2016.02.011.

658 Ellis K, Solar-Lezama A, Tenenbaum JB (2015) Unsupervised Learning by Program
659   Synthesis. In: NIPS'15 Proceedings of the 28th International Conference on Neural
660   Information Processing Systems, pp 973–981. Montreal.

661 Fleuret F, Li T, Dubout C, Wampler EK, Yantis S, Geman D (2011) Comparing machines

662   and humans on a visual categorization test. Proc Natl Acad Sci U S A 108:17621–
663   17625.

664 Göldi M, van Poppel EAM, Rasch B, Schreiner T (2019) Increased neuronal signatures of
665   targeted memory reactivation during slow-wave up states. Sci Rep 9:2715.

666 Grandchamp R, Delorme A (2011) Single-trial normalization for event-related spectral
667   decomposition reduces sensitivity to noisy trials. Front Psychol 2.

668 Groch S, Preiss A, Mcmakin DL, Walitza X, Huber R, Wilhelm I (2017) Targeted Reactivation
669   during Sleep Differentially Affects Negative Memories in Socially Anxious and Healthy
670   Children and Adolescents. 37:2425–2434.

671 Hennies N, Lambon Ralph MA, Durrant SJ, Cousins JN, Lewis PA (2017) Cued memory
672   reactivation during sws abolishes the beneficial effect of sleep on abstraction. Sleep 40.

673 Hu X, Antony JW, Creery JD, Vargas IM, Bodenhausen G V, Paller KA (2015) Unlearning
674   implicit social biases during sleep. Science (80-) 348:1013–1015.

675 Hu X, Cheng LY, Chiu MH, Paller K (2019) Promoting memory consolidation during sleep : A
676   meta-analysis of targeted memory reactivation. Psychol Bull.

677 Hu X, Cheng LY, Chiu MH, Paller KA (2020) Promoting memory consolidation during sleep:
678   A meta-analysis of targeted memory reactivation. Psychol Bull.

679 Hutchison IC, Pezzoli S, Tsimpanouli ME, Abdellahi MEA, Lewis PA (2021) Targeted
680   memory reactivation in REM but not SWS selectively reduces arousal responses.
681   Commun Biol 4.

682 Ibáñez AM, San Martín R, Hurtado E, López V (2009) ERPs studies of cognitive processing
683   during sleep. Int J Psychol 44:290–304.

684 Kirov R, Kolev V, Verleger R, Yordanova J (2015) Labile sleep promotes awareness of
685   abstract knowledge in a serial reaction time task. Front Psychol 6.

686 Landmann N, Kuhn M, Maier J-G, Spiegelhalder K, Baglioni C, Frase L, Riemann D, Sterr A,
687   Nissen C (2015) REM sleep and memory reorganization: potential relevance for
688   psychiatry and psychotherapy. Neurobiol Learn Mem Available at:
689   http://www.ncbi.nlm.nih.gov/pubmed/25602929 [Accessed January 26, 2015].

690 Lau H, Tucker M a., Fishbein W (2010) Daytime napping: Effects on human direct
691   associative and relational memory. Neurobiol Learn Mem 93:554–560.

692 Lehmann M, Schreiner T, Seifritz E, Rasch B (2016) Emotional arousal modulates oscillatory
693   correlates of targeted memory reactivation during NREM, but not REM sleep. Sci Rep
694   6:39229.

695 Lerner I, Gluck MA (2019) Sleep and the extraction of hidden regularities: A systematic
696   review and the importance of temporal rules. Sleep Med Rev 47:39–50.

697 Lewis P a, Durrant SJ (2011) Overlapping memory replay during sleep builds cognitive
698   schemata. Trends Cogn Sci 15:343–351 Available at:
699   http://www.ncbi.nlm.nih.gov/pubmed/21764357 [Accessed December 10, 2014].

700 Lewis PA, Knoblich G, Poe G (2018) How Memory Replay in Sleep Boosts Creative
701   Problem-Solving. Trends Cogn Sci 22:491–503 Available at:
702   http://dx.doi.org/10.1016/j.tics.2018.03.009.

703 Louie K, Wilson MA (2001) Temporally structured replay of awake hippocampal ensemble
704   activity during rapid eye movement sleep. Neuron 29:145–156.

Lundqvist D, Flykt A, Ohman A (1998) The Karolinska directed emotional faces (KDEF). CD ROM from Dep Clin Neurosci Psychol Sect Karolinska Institutet:91–630.

Lutz ND, Diekelmann S, Hinse-Stern P, Born J, Rauss K (2017) Sleep Supports the Slow Abstraction of Gist from Visual Perceptual Memories. Sci Rep 7:42950.

Mainieri G, Maranci JB, Cousyn L, Arnulf I (2019) Overt replay during REM sleep of the UPDRS finger taps sequence in a patient with REM sleep behavior disorder. Sleep Med 63:38–40.

Maquet P, Laureys S, Peigneux P, Fuchs S, Petiau C, Phillips C, Aerts J, Del Fiore G, Degueldre C, Meulemans T, Luxen A, Franck G, Van Der Linden M, Smith C, Cleeremans A (2000) Experience-dependent changes in changes in cerebral activation during human REM sleep. Nat Neurosci 3:831–836.

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164:177–190.

Molle M, Marshall L, Gais S, Born J (2002) Grouping of spindle activity during slow oscillations in human non-REM sleep. J Neurosci 22:10941–10947 Available at: http://www.jneurosci.org/cgi/content/abstract/22/24/10941.

Oostenveld R, Fries P, Maris E, Schoffelen J-M (2011) FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011:156869.

Otero TM (2017) Brief review of fluid reasoning: Conceptualization, neurobasis, and applications. Appl Neuropsychol Child 6:204–211.

Oudiette D, Paller KA (2013) Upgrading the sleeping brain with targeted memory reactivation. Trends Cogn Sci 17:142–149.

Pereira SIR, Lewis PA (2020) The differing roles of NREM and REM sleep in the slow enhancement of skills and schemas. Curr Opin Physiol 15:82–88 Available at: https://doi.org/10.1016/j.cophys.2019.12.005.

Peyrache A, Khamassi M, Benchenane K, Wiener SI, Battaglia FP (2009) Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. NatNeurosci 12:919–926.

Picard-Deland C, Aumont T, Samson-Richer A, Paquette T, Nielsen T (2021) Whole-body procedural learning benefits from targeted memory reactivation in REM sleep and task-related dreaming. Neurobiol Learn Mem 183:107460.

Picton TW (1992) The P300 wave of the human event-related potential. J Clin Neurophysiol 9:456–479.

Rakowska M, Abdellahi MEA, Bagrowska Pa, Navarrete M, Lewis PA (2021) Long term effects of cueing procedural memory reactivation during NREM sleep. Neuroimage 244:118573.

Rasch B, Born J (2013) About sleep's role in memory. Physiol Rev 93:681–766.

Rasch B, Büchel C, Gais S, Born J (2007) Odor cues during slow-wave sleep prompt declarative memory consolidation. Science 315:1426–1429.

Redondo RL, Morris RGM (2011) Making memories last: the synaptic tagging and capture hypothesis. Nat Neurosci Rev 12:17–30.

Rihm JS, Rasch B (2015) Replay of conditioned stimuli during late REM and stage N2 sleep influences affective tone rather than emotional memory strength. Neurobiol Learn Mem

749   122:142–151.

750   Rudoy JD, Voss JL, Westerberg CE, Paller KA (2009) Strengthening individual memories by
751   reactivating them during sleep. Science (80-) 326:1079.

752   Schouten DI, Pereira IR, Tops M, Louzada FM (2017) State of the art on targeted memory
753   reactivation : Sleep your way to enhanced cognition. 32:123–131.

754   Seibt J, Frank MG (2019) Primed to Sleep: The Dynamics of Synaptic Plasticity Across Brain
755   States. Front Syst Neurosci 13:2.

756   Shanahan LK, Gjorgieva E, Paller KA, Kahnt T, Gottfried JA (2018) Odor-evoked category
757   reactivation in human ventromedial prefrontal cortex during sleep promotes memory
758   consolidation. Elife:1–21 Available at: https://elifesciences.org/articles/39681.

759   Simon KCNS, Gómez RL, Nadel L (2018) Losing memories during sleep after targeted
760   memory reactivation. Neurobiol Learn Mem 151:10–17.

761   Sterpenich V, Schmidt C, Albouy G, Matarazzo L, Vanhaudenhuyse A, Boveroux P,
762   Degueldre C, Leclercq Y, Balteau E, Collette F, Luxen A, Phillips C, Maquet P (2014)
763   Memory reactivation during rapid eye movement sleep promotes its generalization and
764   integration in cortical stores. Sleep 37:1061–1075, 1075A-1075B.

765   Tamminen J, Lambon Ralph MA, Lewis PA (2017) Targeted memory reactivation of newly
766   learned words during sleep triggers REM-mediated integration of new memories and
767   existing knowledge. Neurobiol Learn Mem 137:77–82.

768   Wagner U, Gais S, Haider H, Verleger R, Born J (2004) Sleep inspires insight. Nature
769   427:352–355.

770   Welling H (2007) Four Mental Operations in Creative Cognition: The Importance of
771   Abstraction. Creat Res J 19:163–177.

772   Wilhelm I, Rose M, Imhof KI, Rasch B, Buchel C, Born J, Büchel C (2013) The sleeping child
773   outplays the adult's capacity to convert implicit into explicit knowledge. NatNeurosci
774   16:391–393 Available at: http://www.ncbi.nlm.nih.gov/pubmed/23434910.

775   Winkler I, Denham S, Escera C (n.d.) Auditory Event-related Potentials.

776   Witkowski S, Noh S, Lee V, Grimaldi D, Preston AR, Paller KA (2021) Does memory
777   reactivation during sleep support generalization at the cost of memory specifics?
778   Neurobiol Learn Mem 182:107442 Available at:
779   https://doi.org/10.1016/j.nlm.2021.107442.

780
781

782
783
784
785
786

Table 1. Sleep architecture (n = 27)

Sleep variable	Mean	SEM
TST (min)	490.3	10.5
Sleep latency (min)	20.6	2.9
WASO (min)	15.9	3.8
Micro-arousals (#)	39.3	4.6
NREM 1 (min)	33.5	3.0
NREM 2 (min)	254.9	7.9
SWS (min)	85.7	4.2
REM (min)	100.3	4.9
WASO (%)	3.2	3.2
NREM 1 (%)	6.7	0.5
NREM 2 (%)	52.0	1.2
SWS (%)	17.7	0.9
REM (%)	20.3	0.9

787   Total sleep time (TST); Wake after sleep onset (WASO).

788
789

795

Table 2 – Correlations between Cueing Benefit* in REM and SWS

			Pearson's *r*	p†
SWS Pre-sleep	with	REM Pre-sleep	-0.205	0.304
SWS Day 1	with	REM Day1	-0.003	0.987
SWS Day 7	with	REM Day7	-0.147	0.465
SWS Overnight	with	REM Overnight	-0.086	0.669
SWS Week	with	REM Week	-0.207	0.300
SWS Total	with	REM Total	-0.338	0.085

796

797 *cueing benefit = cued – non-cued; Overnight = Day 1 – Pre-sleep; Week = Day 7 – Day 1;
798 Total = Day 7 – Pre-sleep; uncorrected †*p*-value.

799

800

801 **Table 3. Spindles and Slow Oscillations identified in epochs after control and**
802 **experimental sounds**

	Number	Duration (samples)
Spindles		
Control	43.75 (2.01)	78.16 (0.28)
Experimental	43.63 (2.02)	77.20 (0.29)
Slow Oscillations		
Control	59.96 (2.17)	215.46 (1.64)
Experimental	56.39 (2.03)	221.03 (1.65)

803

804 **Table 3:** Spindles and slow oscillations summary, averaged across participants and
805 channels separately for control and experimental epochs. Values within brackets indicates
806 SEM.

807

808 **Table 4: Spindles and slow oscillations summary, averaged across participants and**
809 **channels separately for control and experimental epochs. Values within brackets**
810 **indicates SEM error.**

811

Sleep Stage	Cueing Condition	Oscillation	Pearson's *r*	*p*
SWS	Non-cued	Spindles (#)	0.008	0.97
		Sos (#)	-0.015	0.94
	Cued	Spindles (#)	0.118	0.56
		Sos (#)	0.148	0.46
REM	Non-cued	Spindles (#)	0.324	0.10
		Sos (#)	0.231	0.25
	Cued	Spindles (#)	-0.114	0.57
		Sos (#)	0.016	0.94

812 Slow Oscillations (Sos); Number (#). N = 27

813

814
815

816

817

818

819

820

821

822

823 **Figure 1. Experimental design.** A) Before sleep, participants learned to pair each image (a

824 face or a landscape) with an SVRT problem and its associated sound (Problem-Image

825 Association task). Next, they were trained and tested on the SVRT task, where they had to

826 decide whether or not the test image followed the same rule as the reference image for any

827 given problem, as shown in the upper panel in A. For example, in the problem shown here the

828 rule is: each image contains two identical shapes(Fleuret et al., 2011), see Extended Data

829 Figure 1-1 for another example. Immediately before sleep, participants were probed on their

830 ability to recall which sound (speaker symbols) had been paired to which SVRT problem

831 (Problem-Sound Association task). TMR was applied to different problems during REM and

832 SWS during the night (see B). Finally, participants were retested on the SVRT both next

833 morning (post-sleep day 1) and a week later (post-sleep day 7). B) Representative hypnogram

834 depicting the TMR protocol. During TMR in the night, sounds associated with four problems

835 were replayed in SWS and sounds associated with four other problems were replayed in REM.

836 Control sounds that had not been associated with any problems (new sounds) but instead

837 served as controls for auditory responses were also replayed in both sleep stages. Cueing

838 started with the first instance of SWS and REM and terminated once control and experimental

839 sounds had been presented 28 times each (twice per loop, 14 loops).

840

841 **Extended Data Figure 1-1 SVRT stimuli examples**. Sample images from problem 1 (top

842 panel) and problem 2 (bottom panel), that either follow the rule (on the left) or break the rule

843 (on the right)(Fleuret et al., 2011). For problem 1 the rule is that: each picture contains two

844 identical shapes. The squiggly lines were introduced as distractors (not a part of the rule), to

845 increase the difficulty level. For problem 2 the rule is each image contains two shapes of

846 different sizes, the smaller one inside the larger one, roughly centred. The black filling of the

847 smaller shaped was added in some images as a distractor to increase the difficulty level. Other

848 problems had rules relating, for example, to the number of identical shapes (pairs or triplets),

849 their position (mirrored or translated, touching or not touching, inside or outside one another,

850   aligned or not aligned, etc.) or their arrangement (odd shape in the middle, bigger shape at

851   the edge, etc.).

852
853
854   **Figure 2 –TMR in REM improves rule abstraction.**

855   **A)** SVRT accuracy change overnight (post-sleep day 1 – pre-sleep) and across the week

856   (post-sleep day 7 – post-sleep day 1) is plotted for each sleep stage (SWS and REM) and

857   cueing condition (non-cued and cued). A repeated measures ANOVA revealed a significant

858   sleep stage*cueing condition interaction (p = 0.013) and a simple main-effects analysis

859   showed better performance for problems cued in REM, as compared to problems cued in SWS

860   (p = 0.044). See Extended Data Figure 2-1. **B)** In SWS problems (left), there was no difference

861   between cued and non-cued accuracy in any individual session (p > 0.3). In REM problems

862   (right) there was no difference between cued and non-cued conditions on day 1 (p = 0.550),

863   but at day 7, accuracy was higher on cued compared to non-cued problems (p = 0.002). Mean

864   and SEM are depicted, see also Extended Data Figure 2-2. See Extended Data Figure 2-3 for

865   numbers of trials.

866

867   **Extended Data Figure 2-1a – SVRT accuracy at baseline (pre-sleep).** ANOVA with Cueing

868   (cued/non-cued) and Sleep stage (REM/SWS) as factors.

869   **Extended Data Figure 2-1b – TMR benefit.** Repeated measures ANOVA on retention interval

870   (overnight/week) and Cueing (cued/non-cued) and Sleep stages (SWS/REM). Shaded areas

871   highlight significant results. Overnight benefit is calculated as the difference between Post-

872   sleep day 1 and pre sleep and the week performance is calculated as the difference between

873   both post sleep sessions (Day 7 – Day 1).

874   **Extended Data Figure 2-1c – TMR benefit post-hoc analysis.** Paired t-test for REM

875   conditions to understand the differences between cued and non-cued problems per session

876   (Post-sleep Day1 and Day 7) and also the cueing benefit overnight (difference between Post

877   sleep Day1 and Pre-sleep), a week after (Post-sleep Day7 vs. Pre-sleep) and also the

878   difference between Day 7 and Pre-sleep.

879   **Extended Data Figure 2-2: Accuracy on the SVRT per group and session**

880   **Extended Data Figure 2-3. Number of trials used per participant and condition**

881

882

883   **Figure 3 – Event-related Potentials at Cz during Targeted Memory Reactivation**. Cz

884   ERPs in SWS (blue top panel) and REM (red bottom panel) elicited by control (new) and

885   experimental (task-related) sounds. The vertical dashed line at 0 indicates cue onset (200ms

886   long). A cluster analysis revealed a significant difference between ERPs in response to control

887   and experimental sound in REM between 228ms and 400ms (cluster corrected *p =0.048).

888   Data are depicted as mean ± SEM (n = 26).

889

890   **Figure 4 – Spatial distribution of channels with a statistically significant difference**

891   **between experimental and control sounds during REM.** Data is displayed as the averaged

892   difference (n=26) between experimental and control sounds ERPs in 20ms time bins. *

893   Indicates the position of a significant channel.   The time-frequency cluster permutation

894   analysis for these data is shown in Extended Data Figure 4-1.

895   **Extended Data Figure 4-1.** Time-frequency cluster permutation analysis. When more than

896   one cluster is present, the lowest p-value was selected. When no clusters are found is indi-

897   cated by (-). No statistically significant clusters were found.

898   **Figure 5: Spindles and slow oscillations evoked by TMR**. Top row shows the average of

899   differences in spindles following experimental and control TMR cues, while the bottom line

900   shows the same for slow oscillations. Durations are shown on the left and count is shown on

901   the right. Blue colours indicating higher spindle duration/count for control than experimental.

902   Significant clusters are highlighted with a white star.