



# No evidence for an effect of explicit relevance instruction on consolidation of associative memories

Serena Reverberi<sup>\*</sup>, Nils Kohn, Guillén Fernández

Radboud University Medical Center, Donders Institute for Brain, Cognition, and Behavior, Nijmegen, the Netherlands

## ARTICLE INFO

### Keywords:

Memory  
Consolidation  
Relevance  
fMRI  
Hippocampus

## ABSTRACT

Newly encoded memories are stabilized over time through a process or a set of processes termed consolidation, which happens preferentially during sleep. However, not all memories profit equally from this offline stabilization. Previous research suggested that one factor, which determines whether a memory will benefit from sleep consolidation, is future relevance. The aim of our current study was to replicate these findings and expand them to investigate their neural underpinnings.

In our experiment, 38 participants learned two sets of object-location associations. The two sets of stimuli were presented to each participant intermixed and in random order. After study, participants performed a baseline retention test and were thereafter instructed that, after a delay containing sleep, they would be tested and rewarded only on one of the two sets of stimuli. This relevance instruction was revoked, however, immediately before the test. Thus, this manipulation affected memory consolidation while having no influence on encoding and retrieval. This retention interval was monitored via actigraphy recordings.

While the study session was purely behavioral, the test session was conducted in an MRI scanner, thus we collected neuroimaging data at retrieval of relevant compared with non-relevant items.

Behaviorally, we found no effect of the relevance manipulation on memory retention, confidence rating, or reaction time. At a neural level, no effect of relevance on memory retrieval-related brain operations was observed. Contrary to our expectations, the relevance manipulation did not result in improved consolidation, nor in improved subsequent performance at retrieval. These findings challenge previously published results and suggest that future relevance as manipulated here may not be sufficient to produce enhanced memory consolidation.

## 1. Introduction

Each day, we encounter a multitude of new facts and make new experiences. However, only a select few of these events will be remembered days after they occurred. What aspects of our daily experiences determine whether events will be preserved in long-term memory or rather forgotten? This is still an open question in memory research. Previous studies suggested that one such aspect could be future relevance: the expected usefulness of given information for the individual's future plans and behaviors. The present study is the first to explore the effect of future relevance on memory consolidation during sleep using functional magnetic resonance imaging. This technique allows us to investigate the neural underpinnings of the observed effect of future relevance on the consolidation process. The results of this investigation can help elucidate the factors that determine the "fate" of

memories.

A wide body of research established the essential role of the medial temporal lobe, particularly of the hippocampus, for the acquisition of novel memories (Squire and Zola-Morgan, 1991; Squire et al., 2004). However, the hippocampus is not necessary for the maintenance of remote memories. This is demonstrated by lesion data, which illustrate a temporal gradient of retrograde amnesia following hippocampal lesion, where the more remote a memory, the more likely it is spared from amnesia (Scoville and Milner, 1957; Kapur and Brooks, 1999). The process or set of processes, through which newly encoded memories are stabilized over time is termed 'systems consolidation'. During this process, memories are thought to become increasingly independent of the hippocampus and increasingly dependent on neocortical representations potentially bound together by medial prefrontal nodes (Frankland and Bontempi, 2005; Takashima et al., 2006).

<sup>\*</sup> Corresponding author. Department of Movement Sciences, KU Leuven, 101 Tervuursevest, 3001, Leuven, Belgium.

E-mail address: [serena.reverberi@kuleuven.be](mailto:serena.reverberi@kuleuven.be) (S. Reverberi).

Systems consolidation happens preferentially during sleep (Saletin and Walker, 2012; Born and Wilhelm, 2012). Indeed, sleep after learning protects against forgetting (Jenkins and Dallenbach, 1924) and against interference from competing memories (Ellenbogen et al., 2006). It was suggested that sleep benefits consolidation by allowing the reactivation of memories, which during wake would interfere with the ongoing processing of external stimuli (Diekelmann and Born, 2007). This memory reactivation may then aid the stabilization of neocortical memory traces, eventually rendering them sufficient for memory retrieval (Born and Wilhelm, 2012). However, not all memories equally profit from this offline stabilization process: indeed, memory consolidation was shown to selectively enhance some memories over others. This selection mechanism was suggested to represent an optimization of the consolidation process, whereby salient or relevant memories are preferentially preserved while irrelevant ones are forgotten (Rasch and Born, 2013).

Sleep consolidation was demonstrated to selectively improve memory for emotional items (Hu et al., 2006; Wagner et al., 2006; Holland and Lewis, 2007; Nishida et al., 2008; Javadi et al., 2011), items that have been explicitly encoded (Robertson et al., 2004; Saletin et al., 2011), and items for which a reward is expected (Fischer and Born, 2009). Another factor that was shown to influence whether a memory will benefit from subsequent consolidation is future relevance (Scullin and McDaniel, 2010; Wilhelm et al., 2011; van Dongen et al., 2012; Diekelmann et al., 2013). Test expectation, which is the participants' knowledge that they would be tested on learned material, improved memory recall performance following sleep but not wake, in two previous studies (Scullin and McDaniel, 2010; Wilhelm et al., 2011). However, while the observed effects were attributed to enhanced memory consolidation for items expected to be retested, test expectancy may instead have unselectively facilitated the consolidation process as a whole. Furthermore, a following investigation using a similar design could not replicate these results, finding instead that test expectation enhanced memory performance equally following sleep or wake (Wamsley et al., 2016). A study by Diekelmann et al. (2013) investigated the effect of sleep consolidation on the ability to remember and perform a planned behaviour and demonstrated that sleep enhanced performance compared to wakefulness. However, a subsequent study using a more naturalistic task did not find significant overall performance differences between the sleep and wake groups (Barner et al., 2019). These studies illustrate the contradictory results of current research regarding the effects of future relevance on sleep-dependent memory consolidation and highlight the need for further elucidation on the subject.

An experiment that was designed to ensure selective manipulation of the consolidation process alone, while having no influence on encoding and retrieval, was conducted by van Dongen et al. (2012). In this study, participants learned two sets of randomly intermixed stimuli. After encoding, the relevance of the stimuli was manipulated through explicit instructions: the participants were told that on the next day they would be tested and rewarded only on one of the two sets of stimuli. Immediately before this second test, however, the relevance instruction was revoked. The results of this study suggested that items instructed to be relevant were significantly better remembered than those in the non-relevant category, but only when encoding was followed by sleep. Therefore, it is possible that sleep consolidation selectively benefits relevant memories. If relevant items are indeed better consolidated than non-relevant ones, then the systems consolidation theory predicts that they will elicit decreased hippocampal and increased neocortical and in particular ventromedial prefrontal activation, compared with items in the non-relevant category (Born and Wilhelm, 2012). This prediction can only be tested with neuroimaging. However, previous studies of the effect of future relevance of newly encoded material on consolidation during sleep were either strictly behavioral (Scullin and McDaniel, 2010; van Dongen et al., 2012) or combined behavioral measures with electrophysiological recordings during sleep (Wilhelm et al., 2011; Diekelmann et al., 2013; Wamsley et al., 2016; Barner et al., 2019), but

none included neuroimaging. Furthermore, as the results of previous literature on the effects of relevance on memory performance are contradictory, replication of these experiments is necessary to verify the reproducibility of the observed effects.

The aim of our present study is to confirm the behavioral findings of van Dongen et al. (2012) and expand them to include fMRI data collected during retrieval of items previously instructed to be relevant, versus items in a second, non-relevant category. In doing so, we try to address the following question: how do experimentally manipulated relevance instructions affect memory consolidation during sleep as assessed during subsequent test performance and brain activation at retrieval?

We employ an experimental design analogous to that of van Dongen et al. (2012), where the relevance manipulation is only in effect during the consolidation period, while not influencing encoding and retrieval. In line with previous findings, we expect that items instructed to be relevant will be better remembered compared to items in a second, non-relevant category at a recall test following a delay of approximately 14 h and 30 min which includes sleep (hypothesis *a1*). We also expect to find a positive correlation between sleep time and improvement in memory recall for items instructed to be relevant, as observed in the study of van Dongen et al. (2012) (hypothesis *a2*). Furthermore, we expect that brain activation during retrieval of successfully retrieved versus forgotten items, irrespective of relevance condition, will reveal higher activation in the general retrieval network (Rugg and Vilberg, 2013) (hypothesis *b1*). This network shows enhanced activity associated with successful memory retrieval and includes the angular gyrus, posterior cingulate cortex, medial prefrontal cortex, and hippocampal formation. Furthermore, only if a behavioral effect is present (hypothesis *a1* is correct), then neuroimaging data should reveal a difference in brain activation during retrieval of items instructed to be relevant, versus items in a second, non-relevant category. In accordance with the systems consolidation theory (Born and Wilhelm, 2012), this difference should consist of decreased hippocampal, and increased medial prefrontal, activation for relevant items (hypothesis *b2*).

## 2. Materials and methods

During the experiment, subjects initially memorized two sets of stimuli randomly intermixed, one of which was explicitly instructed to be relevant after this learning session. Following a delay of approximately 14 h and 30 min, which included a night of sleep, the relevance instruction was revoked and subjects were tested on both sets of stimuli. The relevance manipulation was thus only in effect during the consolidation period. During the learning session we collected behavioral data only. In the test phase, when subjects were tested on their memory retention of the previously learned items, we collected both behavioral and fMRI data.

Data collection and analysis were conducted at the Donders Institute for Brain, Cognition, and Behaviour of Nijmegen, Netherlands. The experiment was approved by the ethical committee of CMO region Arnhem/Nijmegen. Subjects provided written informed consent prior to their participation in this study. The study protocol was pre-registered on Open Science Framework and available at the following web address: <https://osf.io/kb23s>. No changes were made to the planned methods.

**Participants.** Forty healthy subjects (aged 18–35 y) were recruited through the Radboud University Nijmegen online recruitment system. The sample size was based on the previous behavioral study (van Dongen et al., 2012), which found a relevant-irrelevant effect across testing sessions of size  $d \sim 0.43$ . This effect size coupled with an alpha error of 0.05 and power of .80 results in a minimum number of participants of  $N = 35$ . Participants were selected on the basis of the following criteria: between 18 and 35 years of age, with no psychiatric or neurological disorders, no sleep disorders or chronic sleep problems, no night work shifts or irregular working shifts, no chronic use of medicine, and no

recreational drug use. Participants received monetary compensation for their participation.

**Task and procedure.** Schematic representations of the task and procedure are shown below in Fig. 1. The task used was a picture-location association task. The pictures were two sets of 80 images each, which belonged to the categories of *buildings* or *furniture*. The pictures were in color and matched for size and resolution. The screen locations were represented by two arrays containing 6 locations each, one of which displayed these locations in the East-West (EW) direction and the other in the North-South (NS) direction. The two picture categories were randomly assigned to the EW or NS array, so that half of the participants associated *buildings* with the EW array and half with the NS array.

The experiment was conducted in two sessions. The first (learning session) took place in the afternoon (15.00–18.00), whereas the second (test session) took place in the morning (8.30–10.00). This timing was preferred over a 24-h delay as it reduces wake time during the retention interval. The learning session consisted of three cycles of encoding and retrieval, while the test session consisted of a single retrieval phase.

During encoding, the participants were first instructed to try to remember the location of the different pictures shown on screen; then they were shown the correct picture-location associations, in random order. During the retrieval phase, the pictures were again presented in random order. The participants needed to actively retrieve and indicate the correct associated locations through the movement of an MRI-

compatible joystick. During retrieval, the participants were also asked to indicate their confidence for each of their responses.

In the learning session, the participants underwent three cycles of encoding and retrieval. While the order of picture presentation was random in each cycle, the picture-location associations remained constant. The participants were thus shown the correct picture-location associations three times, allowing them to progressively improve their performance. The final retrieval phase of the learning session provided baseline measures of memory performance, confidence ratings, and reaction times for subsequent analysis. The participants then received the explicit relevance instruction: the experimenter informed them, both verbally and in writing, that they would be retested only on one of the two stimulus categories. Participants were randomly assigned to one of the two categories. In this way, the *relevant* condition corresponded to the *buildings* category for half the subjects, and to the *furniture* category for the other half. Subjects were further informed during this relevance instruction that they could receive a monetary bonus for each correctly remembered picture-location association, but for the *relevant* condition only. This learning session was carried out in the dummy MRI laboratory, so only behavioral data were collected at this point.

During the delay period between the two sessions, participants were instructed to wear a wrist-mounted actigraph monitor (ActiGraph, Pensacola, USA). They were also instructed to fill in a validated sleep self-report questionnaire (Pittsburgh Sleep Diary, Monk et al., 1994), which was to be completed in two steps: just before sleep on the evening of the first experimental day, and just after wake on the morning of the second experimental day.

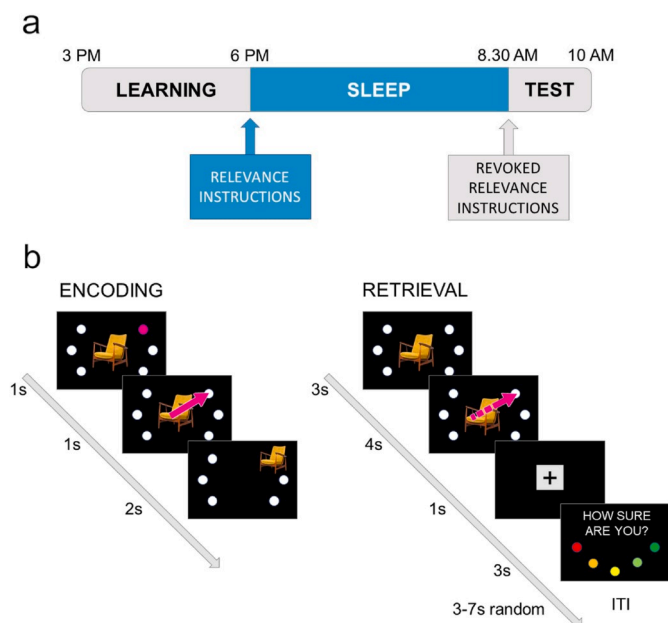
In the second session (test session), the relevance re-instruction was given: the experimenter informed the participants, both verbally and in writing, that they would be retested on both stimulus categories, and that a monetary bonus would be received for each correctly remembered picture-location association, regardless of its relevance condition. The subjects then underwent a single retrieval phase, which was conducted in the MRI scanner. At the end of the experiment, participants completed a questionnaire designed to detect any suspicion of the experimental manipulation and/or any active rehearsal of the picture-location association outside of the two experimental sessions.

**Data collection.** In the learning session only behavioral data were collected. Per trial, the relevant variables were:

1. **Hit or miss.** This was determined by comparing the correct picture location to that chosen by the subject. If the two were equal, the trial was a “hit”, that is the picture-location association was successfully remembered. If the two were not equal, the trial was a “miss”, that is the picture-location association was forgotten. If the participant failed to respond within 7s of the picture presentation, the trial was counted as “null” and discarded from analyses.
2. **Reaction time.** This was calculated by subtracting the time of picture presentation from the time of picture location choice made by the subject. The time of picture location choice was defined as the time the participant’s cursor first reached one of the six locations represented in the location array.
3. **Confidence rating,** as indicated by the subject during the experiment, ranging from rating 1 (not at all confident) to rating 6 (very confident).

In the test session, the same behavioral measures were collected, but in addition fMRI recordings were taken over the whole session. Stimuli were projected onto a mirror attached to the head coil. During the scanning phase, the participant’s eyes were monitored via an eye tracker system (iView X version 2.8.26) to ensure that participants attended to all stimuli.

**fMRI acquisition and preprocessing.** For fMRI data acquisition, a Skyra 3 T MRI scanner (Siemens) was used. An Echo-Planar Imaging sequence was used to acquire whole-brain T2\* -weighted images (66 slices, multi-band-factor = 6; repetition time (TR) = 1 s, echo time (TE) = 35.2 ms, 2



**Fig. 1.** (a) Schematic representation of the experimental design. Participants learn an object-location association task during the afternoon (learning). They are informed that on the following day they will be tested and rewarded only on one set of stimuli (relevance instruction). Following a delay containing sleep, participants are informed that, against expectations, they will be tested on all stimuli (relevance re-instruction). The test session is conducted in the fMRI scanner. The blue color indicates the period during which the relevance manipulation is in effect. (b) Schematic representation of the object-location association task. The task is divided into encoding and retrieval phases. During the encoding phase, participants passively view stimuli moving towards one of six locations on the screen. During the retrieval phase, participants are presented with the stimuli and are instructed to place them in the correct location through movement of a joystick. They are then asked to rate how confident they are with their responses. During the first experimental session (learning session), participants repeat encoding and retrieval phases three times. During the second experimental session (test session) participants complete a single retrieval phase. The inter-trial interval (ITI) is random and ranges 3–7s. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

mm isotropic, flip angle = 60°, slice thickness = 2 mm, field of view (FOV) = 213 mm, interleaved acquisition, Orientation T > C-19.3). A 3D magnetization-prepared rapid gradient echo (MPRAGE) anatomical T1-weighted image (192 slices, 1.0 mm isotropic, TR = 2300 ms, TE = 3.03 ms, flip angle = 8°, slice thickness = 1 mm, FOV = 256 mm, sequential acquisition, sagittal orientation) with whole-brain coverage was acquired for normalization of the functional scans to standard space.

Functional and anatomical images were preprocessed and analyzed using the FSL software (Jenkinson et al., 2012). Functional images were spatially smoothed with a 6 mm full-width half-maximum Gaussian kernel to reduce inter-subject variability. Further preprocessing steps included three-dimensional movement correction. Furthermore, ICA-AROMA was used for denoising (Pruijm et al., 2015). By linear regression of ICA components, AROMA can identify components as noise and (non-aggressively) regresses out the time courses of these components. Subsequently, a high pass filter of 100s was applied. Prior to group analyses, individual functional images were normalized to MNI (MNI ICBM, 2019) space in a two-step procedure combining linear and non-linear registration.

The regressors used for GLM-based analysis were: relevant stimulus hit, relevant stimulus miss, non-relevant stimulus hit, non-relevant stimulus miss and confidence rating presentation, plus additional regressors of no interest such as joystick motion. The standard time-courses of each regressor were convolved with the hemodynamic response function (HRF) in order to predict task-modulated blood oxygen level-dependent (BOLD) response. A nuisance regressor was also used for subjects whose data were corrupted by artifacts only on a subset of the total volumes. The corrupted volumes were manually selected through visual inspection. Convolution with HRF was not applied to this nuisance regressor.

**Data analysis.** The primary outcome variable was the hit ratio (number of correct responses/total number of items) in the first and second sessions. The hit ratio was preferred to the simple number of hits because a few answers were discarded as the participant failed to respond in time. Therefore, the total number of items slightly differed between subjects.

Hit ratios were analyzed with a two-way repeated-measures ANOVA with factors *session* (learning/test) and *relevance* (relevant/non-relevant). This analysis allowed us to test hypothesis a1: higher memory retention for items instructed to be relevant. Reaction times and confidence ratings were analyzed with a 2 × 2 × 2 ANOVA with the factors described above plus the factor *memory* (hit/miss). Hit ratios and reaction times were further analyzed using paired-data Wilcoxon test, as they were not normally distributed. Furthermore, a Bayesian ANOVA was performed on the hit ratios, to provide an estimate of strength of evidence for or against the null hypothesis.

Sleep duration was measured by the actigraph monitor and calculated via the Cole-Kripke algorithm (Cole et al., 1992) as implemented in the R package *actigraph.sleep*. The sleep self-report questionnaire entries for “Went to bed last night at” and “Finally woke at” were used as cut-off times for the actigraphy recordings. The calculated sleep time was investigated to verify whether the total amount of sleep had any effect on memory performance, by performing Pearson correlations between sleep time and hit ratio (number of correct answers/total number of items) at test, and between sleep time and hit ratio at test/hit ratio at learning. This allowed us to test hypothesis a2: positive correlation between sleep time and improvement in memory recall for items instructed to be relevant. All analyses were performed using IBM SPSS Statistics for Windows, version 23 (IBM Corp. Released, 2015), except for Bayesian ANOVA which was performed with JASP 0.9.0.2 (JASP Team, 2019).

The fMRI analysis aimed to reveal the effect of our relevance manipulation on brain areas associated with memory retrieval. To this end, for group analyses a factorial design was used, consisting of a two-way repeated-measures ANOVA using the factors *relevance* (relevant/non-relevant) and *memory* (hit/miss). The main effect of *memory* was investigated to test our hypothesis b1: higher neural activation for

successfully remembered vs forgotten items in the general retrieval network (Rugg and Vilberg, 2013). The *relevance* × *memory* interaction was used to test hypothesis b2: decreased hippocampal and increased medial prefrontal neural activation for remembered items belonging to the relevant, compared with the non-relevant, category. The main effect of *relevance* was also investigated. For each of these contrasts, both the group mean and the covariance with behavioral effect size were calculated. For group statistics, the contrasts of interest were analyzed using the mixed-effects approach FLAME 1 on the whole brain. All z-statistics images for contrasts of interest were thresholded using clusters determined by a cluster forming threshold of  $z > 3.1$  and cluster-level correction at  $p < .05$  for the whole brain (in accordance with Worsley, 2001).

A supplementary analysis was conducted to determine whether any brain region modulated by *memory* displayed activation that scaled with memory confidence as measured by the confidence ratings. For this analysis, new regressors were added to the first-level design for: relevant stimulus hit, relevant stimulus miss, non-relevant stimulus hit, non-relevant stimulus miss. These regressors were each weighted by the normalized confidence rating per trial. Group-level statistics for this analysis were performed with the same procedure detailed above. All fMRI analyses were performed using FSL (6.0.0).

**Data exclusions.** Participants were excluded from any further analysis if they showed suspicion of the experimental manipulation, active rehearsal of the picture-location associations outside of the two experimental sessions, or non-compliance to experimental instructions. Participants were also excluded if they performed at chance level during either of the two experimental sessions. Chance level performance corresponded to a number of hits equal to 27 (160 pictures/6 possible locations).

### 2.1. Comparison with van Dongen et al. (2012)

While the current study closely follows the experimental methods of van Dongen et al. (2012), it is not an identical replication. The following methodological differences should be considered when comparing the two experiments.

**Groups.** The study of van Dongen et al. (2012) included two experimental groups: one was allowed to sleep after learning while the other remained awake. As memory improvement for items instructed to be relevant was observed only in the sleep group, we considered the inclusion of a wake group to be unnecessary for the current study.

**Sample size.** van Dongen et al. (2012) had a total sample size of 50 participants, of which 25 were assigned to the sleep group. We aimed for an increased sample size (sleep group only) of 40 participants. We selected our sample size based on the effect size reported by van Dongen and coworkers, making our study an appropriate replication.

**Number of stimuli.** In the previous study (van Dongen et al., 2012), the number of picture-location associations to be learned was 120, while in the current study it was 160. Participants reached close to 100% recall accuracy in the original study. Consequently, we used more stimuli to increase task difficulty and avoid ceiling effects.

**Instructions.** The instructions given during the two studies were identical. However, two additional checks were performed in the current study, which were not present in the previous one (van Dongen et al., 2012). During the test session, prior to revoking the relevance instructions, the experimenter verbally confirmed that participants remembered which stimulus category had been assigned relevance. This check aimed at ensuring that participants had successfully tagged the correct stimulus category as relevant after the learning session. Additionally, during the debriefing questionnaire administered at the end of the experiment, participants were asked whether they actively rehearsed the picture-location associations instructed to be relevant while outside of the laboratory. Subjects who declared active rehearsal were excluded from all further analyses. This check was performed to minimize the effect of active, conscious rehearsal of the stimuli on

memory consolidation processes.

**Imaging.** While the original study collected exclusively behavioral data, in the current study the test session was conducted in the fMRI scanner. Aside from providing novel neuroimaging data, conducting the test session in the scanner also modifies the participants' study experience in two ways: first, naïve participants may experience novelty or arousal while in the scanner; second, reaction times may be slower when performing the task in the scanner, due to the participants' supine position increasing difficulty to use the joystick. To counteract these possible changes, the learning session was conducted in the dummy scanner. This allowed naïve participants to habituate to the scanning environment, and provided a similar context for learning and test sessions.

### 3. Results

#### 3.1. Experimental subjects

Recruited participants ( $N = 40$ , 30 females, 4 left-handed) had a mean age of  $23.6 \pm 3.2$  years (mean  $\pm$  SD). One participant was excluded from further analysis as they suspected to be tested on non-relevant items, rather than relevant items as instructed. One more participant was excluded due to non-compliance with experimental instructions, resulting in a sample size of 38 subjects for behavioral analysis. Three participants were further excluded from fMRI, but not from behavioral analysis because of technical issues ( $n = 1$ ) or because of lack of "miss" trials (incorrect responses) for one or both of the relevance categories ( $n = 2$ ).

#### 3.2. Behavioral performance

The behavioral analysis mainly aimed at revealing whether the effect of our relevance manipulation resulted in increased memory retention following the delay, increased confidence ratings, or decreased reaction times for relevant compared with non-relevant items (factor *relevance*). Secondary analyses were conducted to evaluate the effect on memory retention, confidence ratings and reaction times of the time interval between the two sessions (factor *session*). We also evaluated whether confidence ratings and reaction times differed between correct and wrong answers (factor *memory*).

**Memory retention.** To determine whether our relevance manipulation resulted in increased memory performance for items instructed to be relevant, compared with those in the non-relevant category, we used the hit ratio (number of correct answers/total number of items) at learning and at test as our primary outcome measures. The group-averaged hit ratios for each condition are reported in Table 1.

A 2-by-2 repeated-measures ANOVA was conducted with "hit ratio" as measure and using within-subject factors *session* (learning/test) and *relevance* (relevant/non-relevant). This test revealed a main effect of *session*,  $F(1,37) = 52.58$ ,  $p < .001$ , but no significant main effect of *relevance* or interaction *session*  $\times$  *relevance*. ANOVA results are summarized in Table 2.

**Table 1**

Group-averaged hit ratios, confidence ratings, and reaction times at learning and test, for relevant and non-relevant items. The values represent the mean  $\pm$  standard deviation.

	Learning		Test	
	Relevant	Non-relevant	Relevant	Non-relevant
<b>Hit ratio</b>	0.71 $\pm$ 0.20	0.71 $\pm$ 0.20	0.67 $\pm$ 0.20	0.66 $\pm$ 0.21
<b>Confidence rating (1–6)</b>				
Hit	4.17 $\pm$ 0.45	4.18 $\pm$ 0.41	3.89 $\pm$ 0.51	3.86 $\pm$ 0.51
Miss	2.31 $\pm$ 0.57	2.44 $\pm$ 0.55	2.20 $\pm$ 0.61	2.13 $\pm$ 0.58
<b>Reaction time (ms)</b>				
Hit	2244 $\pm$ 623	2294 $\pm$ 604	2501 $\pm$ 590	2527 $\pm$ 624
Miss	2878 $\pm$ 895	2915 $\pm$ 803	3123 $\pm$ 882	3140 $\pm$ 861

However, hit ratios were not normally distributed (Shapiro-Wilk test,  $p = .039$ ) in relevant trials. Thus, ANOVA results may not be accurate. However, the non-parametric Wilcoxon paired samples test also shows no significant performance difference ( $p = .479$ ) between relevant and non-relevant trials, consistent with the lack of a main effect of *relevance* detected by ANOVA.

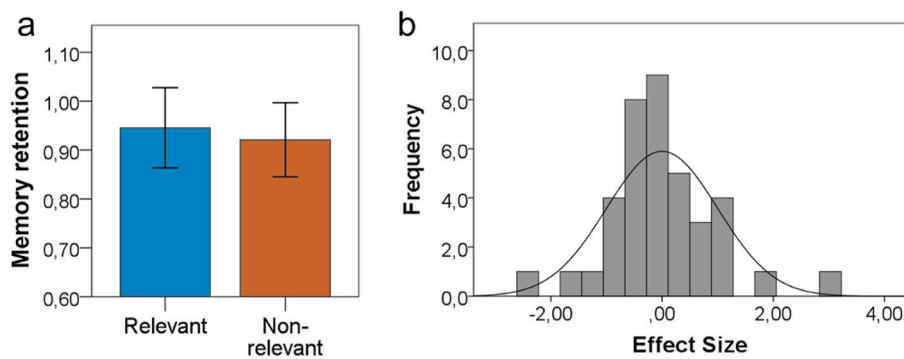
Post-hoc analyses revealed that the main effect of the factor of *session* was driven by an overall forgetting across the test sessions, with participants remembering on average 113.0 items ( $SD = 30.9$ ) at learning, and 106.3 items ( $SD = 32.2$ ) at test. Memory retention, measured as hit ratio during the test session divided by hit ratio for the corresponding category during the learning session, was overall slightly higher for items in the relevant ( $M = 0.95$ ,  $SD = 0.08$ ) compared with non-relevant category ( $M = 0.92$ ,  $SD = 0.07$ ), as shown in Fig. 2a. In absolute number of items, across the two sessions participants forgot on average 2.9 items ( $SD = 3.8$ ) belonging to the relevant category, and 3.8 items ( $SD = 3.3$ ) belonging to the non-relevant category. However, considering the large standard deviations, the mean difference between the two categories is not meaningful. A histogram representing subjects' performances in terms of normalized effect sizes, measured as difference in memory retention between the relevant and non-relevant categories, is presented in Fig. 2b. Normalization was performed by subtracting the group-averaged effect size from each subject's effect size, and dividing by the group standard deviation. The majority of subjects display a near-zero effect size. The non-zero mean on the group level indicating a tendency towards a memory benefit for relevant, compared with non-relevant items appears to be driven by the performance of a few subjects displaying strongly positive effect sizes. To obtain a more robust indication of whether the null hypothesis (no effect of relevance manipulation on memory performance) was to be accepted, we performed a Bayesian repeated-measures ANOVA. The *relevance* model was associated with a Bayes factor  $BF_{01} = 5.53$ . Thus, the observed data is approximately 5 times more likely to be explained by the null than by the alternative hypothesis and constitutes moderate evidence for  $H_0$ .

**Confidence rating.** Confidence ratings were collected per trial and could range from rating 1 (not at all confident) to rating 6 (very confident). The average ratings of each participant were separated by *session* and *relevance*. The group-averaged confidence ratings for each condition are reported in Table 1. Moreover, correct answers were separated from wrong ones (factor *memory*). Data met the assumptions for parametric statistical tests. On average, confidence ratings in the test session only very slightly differed between items in the relevant and non-relevant categories. Hit trials received average ratings of 3.88 ( $SD = 0.51$ ) for

**Table 2**

Summary of results of the analysis of variance. Analyzed measures are presented in bold. Significant results are denoted by \*.

Effect	p
<b>Hit ratio</b>	
<i>session</i>	<.001*
<i>relevance</i>	.802
<i>session</i> $\times$ <i>relevance</i>	.234
<b>Confidence rating</b>	
<i>session</i>	<.001*
<i>memory</i>	<.001*
<i>relevance</i>	.560
<i>session</i> $\times$ <i>relevance</i>	.022*
<i>session</i> $\times$ <i>memory</i>	.178
<i>memory</i> $\times$ <i>relevance</i>	.249
<i>session</i> $\times$ <i>memory</i> $\times$ <i>relevance</i>	.075
<b>Reaction time</b>	
<i>session</i>	.018*
<i>memory</i>	<.001*
<i>relevance</i>	.414
<i>session</i> $\times$ <i>relevance</i>	.478
<i>session</i> $\times$ <i>memory</i>	.886
<i>memory</i> $\times$ <i>relevance</i>	.902
<i>session</i> $\times$ <i>memory</i> $\times$ <i>relevance</i>	.795



**Fig. 2.** (a) Memory retention across test sessions for the relevant and non-relevant categories. Memory retention was measured as the hit ratio during the test session divided by hit ratio for the corresponding category during the learning session. Error bars represent  $\pm 1$  SD. (b) Distribution of subjects' normalized effect sizes, measured as difference in memory retention between the relevant and non-relevant categories across the two experimental sessions. Number of bins = 18. A normal distribution curve is presented in overlay.

relevant, and 3.86 ( $SD = 0.51$ ) for non-relevant items. Miss trials received average ratings of 2.20 ( $SD = 0.61$ ) for relevant, and 2.13 ( $SD = 0.57$ ) for non-relevant items.

A  $2 \times 2 \times 2$  repeated-measures ANOVA was conducted on the measure "confidence rating", using within-subject factors *session* (learning/test), *relevance* (relevant/non-relevant) and *memory* (hits/misses). Significant main effects were observed for factors *session*,  $F(1,35) = 19.92$ ,  $p < .001$ , and *memory*,  $F(1,35) = 298.51$ ,  $p < .001$ , but not for the factor *relevance*. Post-hoc analyses determined that the main effect of *session* consisted in overall lower ratings at test ( $M = 2.99$ ,  $SE = 0.06$ ) compared with learning ( $M = 3.26$ ,  $SE = 0.07$ ). The main effect of *memory* was unsurprisingly a result of higher confidence ratings for hits ( $M = 3.99$ ,  $SE = 0.07$ ) compared with misses ( $M = 2.26$ ,  $SE = 0.08$ ). A significant interaction of *session*  $\times$  *relevance* was observed. This was due to the confidence rating of relevant items being lower than that of non-relevant items in the learning phase (*relevant*:  $M = 3.22$ ,  $SE = 0.06$ ; *non-relevant*:  $M = 3.29$ ,  $SE = 0.06$ ), while the opposite occurred in the test phase (*relevant*:  $M = 3.01$ ,  $SE = 0.07$ ; *non-relevant*:  $M = 2.97$ ,  $SE = 0.07$ ). This interaction is probably spurious as relevance cannot have exerted any effect in the learning phase, and confidence ratings at test did not significantly differ for relevant compared with non-relevant items. No other significant interactions were detected. ANOVA results are summarized in Table 2.

**Reaction time.** Reaction time is the time (in milliseconds) between the presentation of the stimulus and the answer by the participant. The average times for each participant were separated by *session*, *relevance*, and *memory*, similarly to confidence ratings. The group-averaged reaction times for each condition are reported in Table 1. Similar to the results observed for confidence ratings, reaction times during the test session for items in the relevant category (*hits*:  $M = 2501$  ms,  $SD = 590$  ms; *miss*:  $M = 3123$  ms,  $SD = 882$  ms) were comparable to those in the non-relevant category (*hits*:  $M = 2527$  ms,  $SD = 624$  ms; *miss*:  $M = 3140$  ms,  $SD = 861$  ms).

A  $2 \times 2 \times 2$  repeated-measures ANOVA was conducted on the measure "reaction time", using within-subject factors *session* (learning/test), *relevance* (relevant/non-relevant), and *memory* (hits/misses). The results displayed significant main effects of *session*,  $F(1,35) = 6.22$ ,  $p = .018$ , and *memory*,  $F(1,35) = 54.26$ ,  $p < .001$ . Reaction times overall increased across test sessions (*learning*:  $M = 2585$ ,  $SE = 112$ ; *test*:  $M = 2831$ ,  $SE = 113$ ), resulting in the significant main effect of *session*. The main effect of *memory* was driven by lower reaction times for hit trials ( $M = 2411$ ,  $SE = 145$ ) compared with miss trials ( $M = 3005$ ,  $SE = 130$ ). ANOVA results are summarized in Table 2. No significant main effect of *relevance* nor significant interaction with factor of *relevance* with any other factor was observed. However, the distribution of data was not normal in most columns (Shapiro-Wilk test). Thus, we also conducted a paired samples Wilcoxon test for the statistics of interest. Tests confirmed that reaction times were always significantly greater for miss, compared with hit, trials ( $p < .001$  for all contrasts) and during test, compared with learning phase ( $p < .05$  for all contrasts). No main effect of relevance on reaction time was observed using non-parametric test ( $p$

$> .50$ ).

### 3.3. Sleep time

Participants slept on average 6 h and 36 min ( $M = 397$  min,  $SD = 69$ min). van Dongen et al. (2012) reported a positive correlation between sleep time and improvement in memory recall for items instructed to be relevant. However, in the current study no significant correlation was found between sleep time and hit ratio at test, or sleep time and memory retention, for either of the relevance categories.

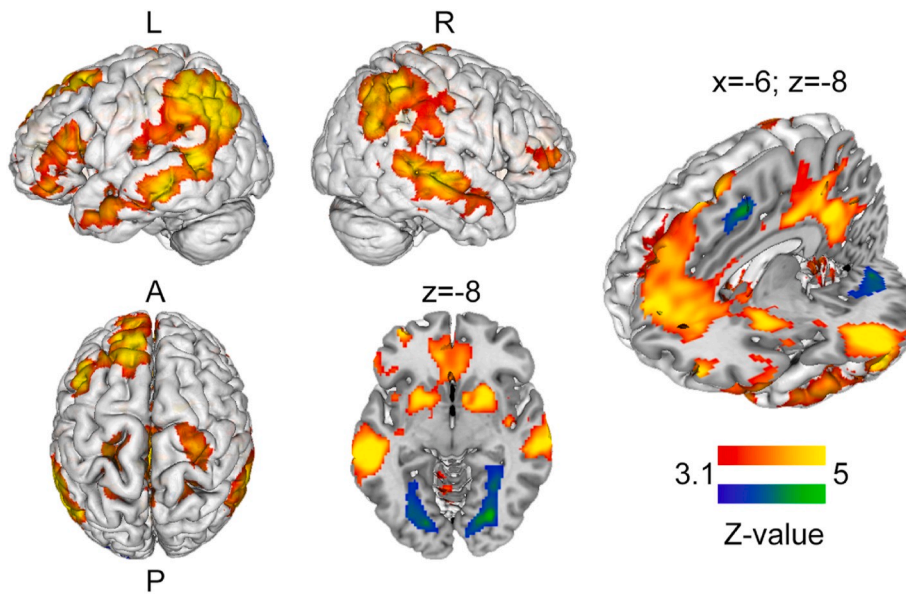
### 3.4. Neuroimaging

The neuroimaging analysis aimed at determining whether our relevance manipulation resulted in different memory consolidation, and consequently different neural activation at retrieval, for relevant and non-relevant items. To this end, we used a factorial design with factors *relevance* (relevant/non-relevant) and *memory* (hit/miss). Furthermore, we endeavored to reveal neural activity directly related to the presence of a relevance effect on the behavioral level. Thus, we performed an analysis of covariance on each of our contrasts of interest, using as covariate the behavioral effect size per subject. The effect size was measured as the difference in memory retention between items in the relevant and non-relevant categories. Finally, we wished to detect modulation of the retrieval process by memory confidence. To this purpose, we performed an analysis of covariance on factors related to the main effect of *memory*, using as covariate the confidence ratings per trial.

**Memory-relevance interaction.** Behavioral results indicated a lack of effect of the relevance manipulation on memory consolidation. To investigate whether the relevance manipulation affected memory consolidation at a neural level, we examined the interaction term *relevance*  $\times$  *memory*. As expected following the behavioral results, at the group level there was no significant activation due to an interaction of *memory* and *relevance*, neither at the standard cluster threshold of  $z > 3.1$ , nor at the more lenient threshold of  $z > 2.3$ .

**The main effect of memory.** Analysis of the main effect of memory revealed a set of brain regions displaying increased activation during hit, compared with miss, trials (Fig. 3). This included a large bilateral cluster extending across the cingulate and paracingulate gyri, middle temporal gyrus, angular gyrus, and putamen. It also included clusters in the right cerebellar crus II, and bilateral inferior frontal gyrus. On the contrary, bilateral clusters in the visual cortex (V2 to V4) extending into the lingual gyrus and posterior parahippocampal gyrus, as well as left occipital pole, left paracingulate gyrus, and right precuneus displayed decreased activation during hits versus miss trials (Fig. 3). Significant clusters and coordinates of their local maxima are reported in Supplement-Table 2.

We examined whether neural activation in any brain region which displayed the main effect of memory would co-vary with the behavioral effect size, measured as difference in memory retention between items in



**Fig. 3.** Brain regions displaying significantly increased activation for hit versus miss trials (red-yellow) or for miss versus hit trials (blue-green) at retrieval, during an object-location association task. Results are presented in MNI152 space. Z-statistics images were thresholded using clusters determined by  $z > 3.1$  and a corrected cluster significance thresholded at  $p < .05$  for the whole brain. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the relevant and non-relevant categories. No brain regions were found where memory-related activity was modulated by behavioral effect size.

**The main effect of relevance.** Analysis of the main effect of relevance did not reveal any voxel with significantly increased or decreased activation for relevant, compared with non-relevant, items. Similarly, no voxel was found whose activity was both modulated by relevance and co-varied with behavioral effect size.

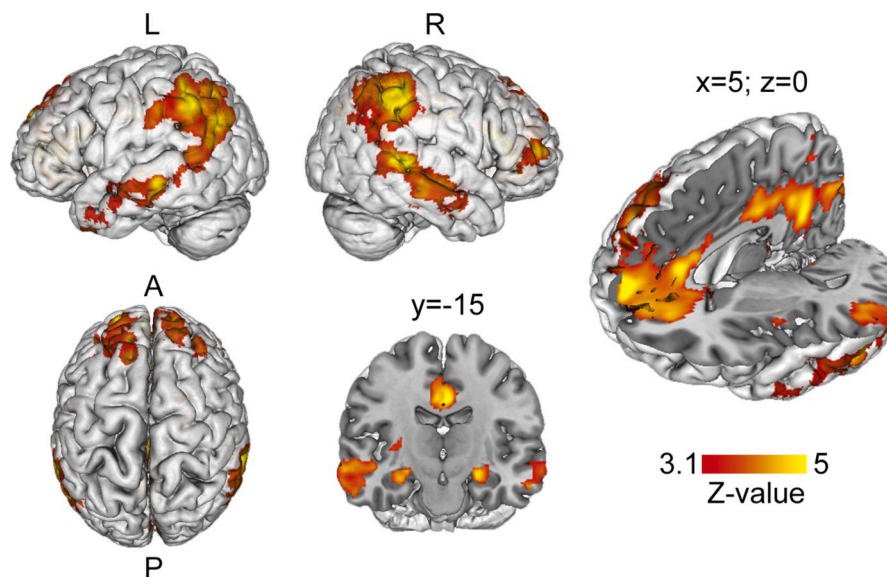
**Modulation of memory retrieval by confidence.** To determine whether neural activity related to the memory effect was modulated by the confidence of responses, we weighed each trial by its normalized confidence rating. One subject was excluded from this analysis due to too few miss trials for the relevant condition, resulting in a total of  $N = 34$  subjects being included in this analysis. When examining the main effect of memory, several brain regions were found to produce increased neural activity with increased confidence ratings (Fig. 4). These included the bilateral cingulate, supramarginal, angular, and superior frontal gyri and hippocampus. The right putamen and insular cortex also displayed memory retrieval activity scaling positively with memory

confidence. Significant clusters and coordinates of their local maxima are reported in Supplement-Table 2.

#### 4. Discussion

Our study does not provide evidence of an effect of explicit relevance instructions as tested here on memory consolidation, nor on memory performance following sleep. On the contrary, the combined behavioral and fMRI results present a convincing argument for the presence of a true null effect.

Memory retention of items belonging to relevant versus non-relevant categories did not significantly differ. Analysis of variance revealed a significant main effect of *session*, but no main effect of *relevance* nor significant interaction of *session* and *relevance*. Furthermore, the effect of relevance instructions on memory appeared to be centered on zero over the whole group, with half of the participants showing a memory benefit for relevant items and the other half a penalty. Bayesian analysis indicated that the observed data constitutes moderate evidence in favor of



**Fig. 4.** Brain regions where neural activity to hit, versus miss, trials were modulated by confidence ratings. Results are presented in MNI152 space. Z-statistics images were thresholded using clusters determined by  $z > 3.1$  and a corrected cluster significance thresholded at  $p < .05$  for the whole brain.

the null hypothesis: relevance did not influence memory retention. Analyses of variance conducted on confidence ratings and reaction times confirmed the lack of a relevance effect on our secondary outcome measures. Taken together, these data suggest that our relevance manipulation did not produce any significant effect at the behavioral level.

Corroborating the behavioral results, fMRI analysis found no neural activation reflecting a *relevance* × *memory* interaction, indicating that the relevance instructions did not affect memory consolidation during sleep. The main effect of *relevance* also produced no results at the standard cluster-forming threshold.

While our study failed to demonstrate a relevance effect on the neural level, the main effect of *memory* indicates that our experimental paradigm led to the successful consolidation of picture-location associations. In fact, the set of brain regions that we found to display increased activation at retrieval of remembered, versus forgotten, items essentially overlaps with the general retrieval network as described in previous literature (Rugg and Vilberg, 2013), which includes the cingulate and angular gyri and the medial frontal cortex. Decreased activation for remembered, versus forgotten, items was instead observed in the hippocampal formation. This could perhaps reflect a form of the repetition suppression effect, whereby familiar (remembered) items produce decreased medial temporal activation compared to novel (forgotten) stimuli (Gonsalves et al., 2005).

We also investigated how memory confidence modulated the memory retrieval process. The brain regions observed in our study where memory confidence modulated the retrieval process were in line with previous literature regarding source and recognition memory. Activity in the angular gyrus and posterior cingulate gyrus was previously shown to reflect source memory effects associated with highly confident judgments (Thakral et al., 2015) while recognition-related activity in the left superior frontal gyrus was found to increase with increasing memory strength (Vilberg and Rugg, 2007). Hippocampal activity during a recognition test was shown to co-vary with confidence of source memory assessments (Yu et al., 2012). A previous recognition memory study further found increased neural activity in the bilateral anterior and posterior cingulate cortex and hippocampus for high-compared with low-confidence responses, regardless of accuracy (Moritz et al., 2006). In addition to these previously identified areas, our results suggest that the supramarginal gyrus and putamen may be involved in the processing of confidence for associative memories.

While some previous studies suggested that relevance instructions increase memory performance following sleep (Scullin and McDaniel, 2010; Wilhelm et al., 2011; van Dongen et al., 2012; Diekelmann et al., 2013), others did not find this effect (Wamsley et al., 2016; Barner et al., 2019). The result of our study is thus in partial agreement with previous literature but differs from those of at least four prior studies. Strikingly, our results differ from the observations of van Dongen et al. (2012) despite having closely followed their task and procedure. This could be due to one of the differences between the two studies (outlined in section 2.1): first of all, van Dongen and colleagues compared effects of relevance instructions on memory performance in both sleep and wake groups, while our design included the sleep group alone. While this did modify the overall composition of the study, our hypotheses were solely based on effects observed in the original study *within* the sleep group. Consequently, the removal of the wake group should not have affected these results. Secondly, our final sample size of 38 participants was larger than the original study's sample size of 25. Therefore, our failure to replicate the previous findings is more likely a result of type I errors in the original study, than of type II in the current one. Furthermore, we increased the number of stimuli to be learned by 30%, due to average near-ceiling performances in the original study. We find that participants' memory performance in our study is at a group level comparable to that reported in the previous study (van Dongen et al., 2012) with an average forgetting across testing sessions of 1 extra item for the non-relevant compared with the relevant category. However, due to

increased total number of items to be learned, the observed variance in our study is higher. The increased variance might effectively have overshadowed any effect of relevance. Finally, we collected fMRI data in addition to behavioral measures and demonstrated the failure of relevance instructions to produce differences in memory retrieval at a neural level as well.

In our study, relevance instructions did not improve memory performance for the items instructed to be remembered. One could speculate that *relevance*, as defined and implemented by us, does not have inherent biological significance. However, the participants were also informed that only relevant items would be monetarily compensated. The anticipated reward should have acted as a significant motivational source. The failure to produce an improved performance for relevant items suggests that our manipulation was not sufficiently strong or salient, or that the task was in general too easy. Another possibility is that repetition of the relevant memories is necessary. Indeed, deliberate retrieval of learned items is known to improve memory consolidation and long-term retention of material (Roediger and Butler, 2011), but it was discouraged in this study. Future research may explore this hypothesis by manipulating both relevance and retrieval frequency of learned material, for example with a study design that employs our relevance manipulation but where repetition is encouraged in one group of participants and discouraged in another. If improved memory retention for relevant items is only observed when repetition is present, our hypothesis would be corroborated.

## 5. Conclusion

In this study, we set out to test the reproducibility of previous findings regarding the effect of relevance instructions on memory consolidation during sleep and subsequent memory retrieval. We employed an associative memory task and used a relevance manipulation designed to selectively affect memory consolidation while not influencing encoding and retrieval processes. Our results indicate that explicit relevance instructions do not produce improved memory consolidation for the items instructed to be relevant. This is true both at a behavioral level, as measured by memory retention, confidence rating, and reaction time, as well as at a neural level, where no brain activity linked to the effect of relevance was observed.

## CRedit authorship contribution statement

**Serena Reverberi:** Investigation, Formal analysis, Writing - original draft, Writing - review & editing, Visualization. **Nils Kohn:** Conceptualization, Methodology, Formal analysis, Writing - review & editing, Supervision. **Guillén Fernández:** Conceptualization, Methodology, Writing - review & editing, Supervision, Project administration.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuropsychologia.2020.107491>.

## References

- Barner, C., Altgassen, M., Born, J., Diekelmann, S., 2019. Effects of sleep on the realization of complex plans. *J. Sleep Res.* 28 (1), e12655.
- Born, J., Wilhelm, I., 2012. System consolidation of memory during sleep. *Psychol. Res.* 76 (2), 192–203.
- Cole, R.J., Kripke, D.F., Gruen, W., Mullaney, D.J., Gillin, J.C., 1992. Automatic sleep/wake identification from wrist activity. *Sleep* 15, 461–469.
- Diekelmann, S., Born, J., 2007. One memory, two ways to consolidate? *Nat. Neurosci.* 10 (9), 1085.
- Diekelmann, S., Wilhelm, I., Wagner, U., Born, J., 2013. Sleep improves prospective remembering by facilitating spontaneous-associative retrieval processes. *PLoS One* 8 (10), e77621.
- Ellenbogen, J.M., Hulbert, J.C., Stickgold, R., Dinges, D.F., Thompson-Schill, S.L., 2006. Interfering with theories of sleep and memory: sleep, declarative memory, and associative interference. *Curr. Biol.* 16 (13), 1290–1294.



- Fischer, S., Born, J., 2009. Anticipated reward enhances offline learning during sleep. *J. Exp. Psychol. Learn. Mem. Cognit.* 35 (6), 1586.
- Frankland, P.W., Bontempi, B., 2005. The organization of recent and remote memories. *Nat. Rev. Neurosci.* 6 (2), 119.
- Gonsalves, B.D., Kahn, I., Curran, T., Norman, K.A., Wagner, A.D., 2005. Memory strength and repetition suppression: multimodal imaging of medial temporal cortical contributions to recognition. *Neuron* 47 (5), 751–761.
- Holland, P., Lewis, P.A., 2007. Emotional memory: selective enhancement by sleep. *Curr. Biol.* 17 (5), R179–R181.
- Hu, P., Stylos-Allan, M., Walker, M.P., 2006. Sleep facilitates consolidation of emotional declarative memory. *Psychol. Sci.* 17 (10), 891–898.
- IBM Corp Released, 2015. IBM SPSS Statistics for Windows, Version 23.0. IBM Corp, Armonk, NY.
- JASP Team, 2019. JASP ([Computer software]), Version 0.9 0.2.
- Javadi, A.H., Walsh, V., Lewis, P.A., 2011. Offline consolidation of procedural skill learning is enhanced by negative emotional content. *Exp. Brain Res.* 208 (4), 507–517.
- Jenkins, J.G., Dallenbach, K.M., 1924. Obliviscence during sleep and waking. *Am. J. Psychol.* 35 (4), 605–612.
- Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. *Neuroimage* 62 (2), 782–790.
- Kapur, N., Brooks, D.J., 1999. Temporally-specific retrograde amnesia in two cases of discrete bilateral hippocampal pathology. *Hippocampus* 9 (3), 247–254.
- MNI ICBM 152 Non-linear 6th Generation Symmetric Average Brain Stereotaxic Registration Model. (2019). Retrieved from: <http://nist.mni.mcgill.ca/?p=858>.
- Monk, T.H., Reynolds, C.F., Kupfer, D.J., Buysse, D.J., Coble, P.A., Hayes, A.J., et al., 1994. The Pittsburgh sleep diary. *J. Sleep Res.* 3 (2), 111–120.
- Moritz, S., Gläscher, J., Sommer, T., Büchel, C., Braus, D.F., 2006. Neural correlates of memory confidence. *Neuroimage* 33 (4), 1188–1193.
- Nishida, M., Pearsall, J., Buckner, R.L., Walker, M.P., 2008. REM sleep, prefrontal theta, and the consolidation of human emotional memory. *Cerebr. Cortex* 19 (5), 1158–1166.
- Pruim, R.H., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J.K., Beckmann, C.F., 2015. ICA-AROMA: a robust ICA-based strategy for removing motion artifacts from fMRI data. *Neuroimage* 112, 267–277.
- Rasch, B., Born, J., 2013. About sleep's role in memory. *Physiol. Rev.* 93 (2), 681–766.
- Robertson, E.M., Pascual-Leone, A., Press, D.Z., 2004. Awareness modifies the skill-learning benefits of sleep. *Curr. Biol.* 14 (3), 208–212.
- Roediger III, H.L., Butler, A.C., 2011. The critical role of retrieval practice in long-term retention. *Trends Cognit. Sci.* 15 (1), 20–27.
- Rugg, M.D., Vilberg, K.L., 2013. Brain networks underlying episodic memory retrieval. *Curr. Opin. Neurobiol.* 23 (2), 255.
- Saletin, J.M., Goldstein, A.N., Walker, M.P., 2011. The role of sleep in directed forgetting and remembering of human memories. *Cerebr. Cortex* 21 (11), 2534–2541.
- Saletin, J.M., Walker, M.P., 2012. Nocturnal mnemonics: sleep and hippocampal memory processing. *Front. Neurol.* 3, 59.
- Scoville, W.B., Milner, B., 1957. Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatr.* 20 (1), 11.
- Scullin, M.K., McDaniel, M.A., 2010. Remembering to execute a goal: sleep on it! *Psychol. Sci.* 21 (7), 1028–1035.
- Squire, L.R., Zola-Morgan, S., 1991. The medial temporal lobe memory system. *Science* 253 (5026), 1380–1386.
- Squire, L.R., Stark, C.E., Clark, R.E., 2004. The medial temporal lobe. *Annu. Rev. Neurosci.* 27, 279–306.
- Takashima, A., Petersson, K.M., Rutters, F., Tendolcar, I., Jensen, O., Zwartz, M.J., et al., 2006. Declarative memory consolidation in humans: a prospective functional magnetic resonance imaging study. *Proc. Natl. Acad. Sci. Unit. States Am.* 103 (3), 756–761.
- Thakral, P.P., Wang, T.H., Rugg, M.D., 2015. Cortical reinstatement and the confidence and accuracy of source memory. *Neuroimage* 109, 118–129.
- van Dongen, E.V., Thielen, J.W., Takashima, A., Barth, M., Fernández, G., 2012. Sleep supports selective retention of associative memories based on relevance for future utilization. *PLoS One* 7 (8), e43426.
- Vilberg, K.L., Rugg, M.D., 2007. Dissociation of the neural correlates of recognition memory according to familiarity, recollection, and amount of recollected information. *Neuropsychologia* 45 (10), 2216–2225.
- Wagner, U., Hallschmid, M., Rasch, B., Born, J., 2006. Brief sleep after learning keeps emotional memories alive for years. *Biol. Psychiatr.* 60 (7), 788–790.
- Wamsley, E.J., Hamilton, K., Graveline, Y., Manceor, S., Parr, E., 2016. Test expectation enhances memory consolidation across both sleep and wake. *PLoS One* 11 (10), e0165141.
- Wilhelm, I., Diekelmann, S., Molzow, I., Ayoub, A., Mölle, M., Born, J., 2011. Sleep selectively enhances memory expected to be of future relevance. *J. Neurosci.* 31 (5), 1563–1569.
- Worsley, K.J., 2001. Statistical analysis of activation images. *Functional MRI: an Introduction to Methods*, pp. 251–270.
- Yu, S.S., Johnson, J.D., Rugg, M.D., 2012. Hippocampal activity during recognition memory co-varies with the accuracy and confidence of source memory judgments. *Hippocampus* 22 (6), 1429–1437.