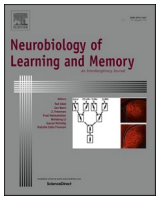
Neurobiology of Learning and Memory 219 (2025) 108049

Contents lists available at ScienceDirect

Neurobiology of Learning and Memory

journal homepage: www.elsevier.com/locate/ynlme



Cortisol modulates hippocampus activation during semantic substitution in men

Nicole Klein, Carina Zöllner, Tobias Otto, Oliver Tobias Wolf, Christian Josef Merz *

Department of Cognitive Psychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr University Bochum, Germany

ARTICLE INFO

Keywords:
Episodic memory
fMRI
Prior knowledge
Recognition
Schema
Scenario construction

ABSTRACT

In the case of incomplete episodic memory retrieval, semantic knowledge may play a vital role compared to random memory errors in filling in memory gaps (semantic substitution). Stress impairs (episodic) memory retrieval via stress hormones (mainly cortisol) targeting the hippocampus. This preregistered neuroimaging study aimed to examine the neural mechanisms of the interplay between episodic memories and prior knowledge during the reconstruction of a past scenario under elevated cortisol levels in men. During encoding, sixty men prepared a virtual apartment for having guests over by using button presses to interact with household objects (e.g., toasting a slice of bread) that were placed congruently to semantic knowledge (e.g., a coffee machine in the kitchen) or incongruently (e.g., a toaster in the bathroom). One day later, participants received (order randomized, double-blind) either 20 mg of cortisol (n = 30) or a placebo (n = 30) before a recognition task. After identifying objects as old, we included a room recall using a forced-choice question in which room the objects were remembered. For incongruent objects this allowed us to differentiate the involvement of episodic, semantic, or random memory. Cortisol did not impair general recognition memory. The manipulation of stimuli during encoding, as being congruent and interactable (relevant to the goal) appears to be predictive of later accurate room recall. Semantic substitution in case of episodic memory failure was associated with anterior parahippocampal and gyrus rectus activation. Cortisol administration increased hippocampal activation during semantic substitution, suggesting a compensatory effect. The results characterized the neural correlates of semantic substitution and speak for an intertwined view of episodic memory and semantic knowledge, which is further shaped by the stress hormone cortisol.

1. Introduction

Remembering enables us to recall memorable events (such as weddings or accidents) or mundane occurrences (where we left our phones) and to learn from past experiences. Misremembering details of such events can lead to conflicts, such as waiting at the wrong place for a meeting or falsely accusing a suspect of wearing glasses, which can thus influence a subsequent investigation. Comprehending the underlying mechanisms that facilitate memory errors when reconstructing past scenarios is crucial.

One factor that may contribute to systematic memory errors during scenario construction is semantic information. For example, new information that is congruent to prior knowledge (organized in a schema) compared to incongruent improves memory (Bein et al., 2015; Brod & Shing, 2019). Importantly, the scenario construction model (SCM; Cheng et al., 2016) offers a valuable understanding of how semantic information is incorporated into episodic memory. The SCM proposes that when reconstructing a past event and only the gist is accessible (left the phone somewhere at home), prior knowledge (usually the phone is on the kitchen counter) is used to fill in missing details (in fact, the phone was left on the bed). Hereafter, we will use the term'semantic substitution' to describe the preference for semantic information over unrelated information (wrong details) in case of incomplete episodic memory retrieval. Previously, we investigated semantic substitution experimentally using a virtual memory paradigm (Zöllner et al., 2022). Participants actively encoded a virtual episode by navigating through a virtual apartment on a computer. By using button presses, they interacted (preparing a slice of bread using a toaster) with some objects that were either placed congruently (a toaster in the kitchen) or incongruently (a toaster in the bathroom) in rooms based on semantic knowledge.

* Corresponding author at: Department of Cognitive Psychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr University Bochum, Universitätsstr. 150, 44780 Bochum, Germany.
E-mail address: Christian.J.Merz@rub.de (C.J. Merz).

https://doi.org/10.1016/j.nlm.2025.108049
Received 30 April 2024; Received in revised form 27 February 2025; Accepted 24 March 2025
Available online 25 March 2025

On the next day, they were asked to recognize the objects and recall in which room they had seen them prior. Results showed that participants were more likely to remember incongruent objects in the room where they belong semantically than in an unrelated room, if they could not recall the episodic room. Similarly, during the report of an autobiographical event, participants tend to use semantic information to supplement missing episodic details and a negative correlation between episodic and semantic details was found (Devitt et al., 2017).

The underlying neural correlates of episodic and semantic memory are relatively well understood individually, but it is still unclear how they jointly contribute to scenario reconstruction. Episodic memory retrieval, for example, is associated with the activation of prefrontal (Preston & Eichenbaum, 2013) and parietal regions, particularly the precuneus (Cavanna & Trimble, 2006) and, most importantly, the hippocampus (HPC; Barry & Maguire, 2019; Sekeres et al., 2018). The parahippocampal gyrus (PHG) can be divided based on its functionality for episodic and semantic memory. The posterior part (pPHG) is associated with recollection (episodic memory), while the anterior part (aPHG) contributes to the feeling of familiarity (semantic memory) (Diana et al., 2007). Furthermore, for semantic memories, the temporal and parietal lobes, particularly the angular gyrus, play crucial roles (Binder & Desai, 2011). Noticeable, episodic and semantic memories also share neural underpinnings (Binder et al., 2009; Rugg & Vilberg, 2013. For instance, the HPC binds different elements of an episode (such as shapes, smells, and colors) that are spread out across various neocortical systems (Moscovitch, 1994). Besides, concept cells suggested to be involved in representing abstract concepts are located in the HPC, among others (Quiroga et al., 2005). Similarly, studies have shown that the medial prefrontal cortex (mPFC), which is also relevant for episodic memories, detects and integrates schema-related information (Brod et al., 2013; Ghosh & Gilboa, 2014; van Kesteren et al., 2010). More specifically, gist-like memories containing high perceptual details activate the HPC, while schema representations activate the ventromedial prefrontal cortex (vmPFC; Robin & Moscovitch, 2017). While these findings may provide initial evidence for the proposed neural interplay, the exact neural correlates involved in scenario construction, particularly semantic substitution, remain unclear. It may be necessary to consider the anterior cingulate cortex (ACC) to better understand why semantic information is preferred over irrelevant information when completing memories. As a conflict monitor (Botvinick et al., 1999), the ACC could be responsible for filtering out irrelevant details, thereby promoting the preference for semantically relevant content.

Importantly, some of these brain structures are sensitive to the effects of stress hormones. Acute stress triggers two routes: The fast sympathetic nervous system (SNS) and the slower-acting hypothalamus-pituitary-adrenocortical (HPA) axis (de Kloet et al., 2005). As end products, the SNS releases the catecholamines (nor)epinephrine from the adrenal medulla, while the HPA axis releases the glucocorticoid cortisol from the adrenal cortex (Lupien et al., 2007). Cortisol then binds to mineralocorticoid and glucocorticoid receptors (de Kloet et al., 2005), thereby modulating neural activation in various brain regions, including HPC and prefrontal cortex (PFC; Dedovic et al., 2009; Herman et al., 2005). Cortisol administration typically impairs memory retrieval (Shields et al., 2017) and was associated with reduced HPC activation in a recognition task (Oei et al., 2007). Elevated cortisol concentrations might thus influence the involvement of semantic memory in creating a past scenario due to impaired episodic contribution and corresponding decreased activity in regions relevant to episodic memory.

In the current study, we attempted to replicate the behavioral findings of semantic substitution from our initial study (Zöllner et al., 2022) as described above, to investigate the underlying neural correlates and how cortisol influences the balance between episodic memories and semantic knowledge. Therefore, participants experienced an episodic event (preparing the home for having a date over) in a virtual home during encoding by interacting with some stimuli, i.e., neutral household objects (preparing a slice of bread using a toaster). These objects were found congruently (toaster in the kitchen) or at odds with prior knowledge (toaster in the bathroom). After receiving cortisol or a placebo the next day, participants underwent a modified recognition task during functional magnetic resonance imaging (fMRI), not only recognizing the objects but also indicating the room in which they remembered seeing them in the virtual apartment using an additional forced-choice question, i.e. room recall (RR). To investigate semantic substitution, only information from the RR of incongruent objects was used, disentangling the involvement of episodic memory, semantic knowledge, or random errors. Semantic substitution does not equal mere semantic retrieval. Instead, semantic substitution involves the preferred integration, or potentially expedited retrieval, of semantic information over random or wrong information to substitute missing details, particularly when episodic memory cannot be fully retrieved during scenario reconstruction. On a behavioral level, we first assumed that participants rely more on semantic information than on unrelated information during retrieval when experiencing episodic memory failure (semantic substitution). In addition, cortisol administration should reduce episodic memory accuracy, resulting in more pronounced semantic substitution.

For our hypothesis on the neural correlates of semantic substitution, we must consider the complexity of the process, which existing findings on semantic retrieval alone cannot fully explain. Instead, we integrated the separate findings of episodic and semantic memory's sensitivity to cortisol and their neural correlates. Therefore, we hypothesized that during semantic substitution, participants in the cortisol group show decreased activity in the HPC, precuneus (episodic memory; Cavanna & Trimble, 2006; Oei et al., 2007), and vmPFC (semantic knowledge; Robin & Moscovitch, 2017), but increased activity in the angular gyrus (semantic knowledge; Binder & Desai, 2011) and ACC (conflict monitoring; Botvinick et al., 1999) compared to the placebo group.

For a deeper understanding of general memory processes during the construction of a past scenario, we explored the differences in neural correlates between episodic and semantic memory retrieval by focusing on findings of cortisol influences on episodic and semantic memory separately. Comparing correct retrieval of episodic content and semantic substitution, we expected cortisol to decrease HPC and posterior precuneus (episodic memory; Cavanna & Trimble, 2006; Oei et al., 2007) activation and increase activation in the angular gyrus (semantic knowledge; Binder & Desai, 2011) and aPHG (feeling of familiarity, Diana et al., 2007) compared to placebo.[1] Finally, we were also interested in how cortisol might modulate brain regions, differentiating between episodic and wrong memories in an exploratory manner.

2. Material and methods

2.1. Participants and general procedure

This sample size was based on low, medium-range effect sizes found in studies using cortisol manipulations before retrieval (Het et al., 2005) for our preregistration (https://osf.io/7fc4s/?view_only). Using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009), we computed the sample size necessary for a repeated-measures ANOVA with the between-subjects factor treatment (cortisol vs placebo) and the within-subjects factor time (one day vs one month; relevant for another research question of the project), a correlation between repeated measures of 0.5 (Zöllner et al., 2022), an effect size of $f^2 = 0.25$, power of $1\text{-}\beta = 0.95$ and a Type I error of $\alpha = 0.05$. A required sample size of 54 was

[1] Please note that we made a correction in our hypotheses from the preregistration. Specifically, we swapped the region ACC (from hypothesis 5) and the aPHG (from hypothesis 4) with one another. This was a result of a mistake on our part, as we inadvertently listed them under the wrong hypothesis in the preregistration. However, we provided the accurate description of their functions.

necessary. Due to the absence of tools and reliable functions to calculate fMRI power, we followed the standard participant numbers from recent studies (e.g. Hagedorn et al., 2021). Thus, relying on behavioral power analysis and insights from prior fMRI research, we increased the sample size to 60, i.e., 30 per group, considering the possibility of dropouts in fMRI studies.

In total, 60 healthy, right-handed men participated in the study. Women were excluded due to the influence of fluctuating sex hormones over the menstrual cycle on memory and neural correlates in interaction with cortisol (Merz & Wolf, 2017) and to gain first insights in a quite homogenous sample, as the study aimed to provide proof of concept. Only healthy, non-smoking individuals between 18 and 38 years who were not sensitive to motion sickness and met standard fMRI criteria were included in the study. In addition, participants were instructed to avoid eating, drinking anything except water, and exercising two hours before each session. The study involved two consecutive afternoon fMRI scans from 1 pm to 6 pm to maintain relatively constant endogenous cortisol levels and an online memory assessment one month later. For this study, we only consider days with scanning. On the first day, participants took a COVID-19 test, signed consent forms, completed two tasks inside the scanner; a picture-viewing task (part of another project, Zöllner et al., subm.*)* and a virtual reality-based encoding task, and filled out the Igroup Presence Questionnaire (IPQ, Schubert et al., 2001) to assess immersion. On the second day, participants completed a demographic questionnaire and a filler task during the 30-minute waiting period after tablet intake. Memory tests occurred in the following order: free recall, recognition task, spatial memory tasks, and temporal memory tasks. Only the recognition task was performed inside the scanner, which is the only task relevant to this manuscript. Participants responded using MRI-compatible button pads during all tasks performed in the scanner on both days. Memory performance was tested again one month later using Qualtrics (Qualtrics, 2005, Provo, Utah, USA, Version 08/2021–10/2022) and an anonymized voice memo cloud upload, which was not analyzed for the current manuscript. Finally, participants were reimbursed with 60€ and received clarification of the study. All procedures conformed to the Declaration of Helsinki and were approved by the ethics committee of the Medical Faculty of the Ruhr University Bochum (reg.-nr.:18–6368).

2.2. Encoding task

During encoding, participants moved through a virtual apartment from a first-person perspective. The right-hand fingers were used for navigation, and the left-hand fingers were used for displaying (index finger) and executing (middle finger) tasks with interactable objects. After a familiarization phase with the controls and the environment, the main episodic virtual environment (EVE) task followed. Each room (kitchen, bathroom, bedroom) in the apartment contained eight objects, resulting in 24 objects. Objects in the rooms were randomized and controlled for congruency and interaction. Half of the objects in each room were placed congruently (*a coffee machine in a kitchen*) or incongruently (*a toaster in the bathroom*) with regard to the semantic relation of the room. Two congruent and two incongruent objects in each room were interaction objects. All tasks, i.e., interaction with 12 specific objects (6 congruent and 6 incongruent), were part of a cover story (*moving into the apartment and preparing everything for having a date coming over*) to create a coherent storyline. The order of the tasks was randomized and followed specific rules to ensure that no interaction with all objects of one semantic category occurred consecutively to prevent biases on memory, such as clustering. No more than two subsequent tasks were allowed to occur in the same room and within the same semantic category. Each task (e.g. *'you are already exhausted from all the tidying, go prepare a toast as a little snack for you'*) required locating and identifying a specific object (e.g. *a toaster*) and holding the execution button for five seconds. Consequently, the object or a related part was lifted with a loading bar indicating progress. The object may change into its final state after completion (see Fig. 1). The next task was displayed after pressing the instruction button. To control for size, 12 small (e.g., ear plugs) and 12 large objects (e.g., a bathing towel) were used based on ratings from an independent sample (Zöllner et al., 2022). Due to the nature of the sample set and to ensure applicability to real-world scenarios objects were balanced by size per room but neither manipulated nor analyzed.

2.3. Retrieval

We used a modified old/new recognition task to assess memory and semantic substitution specifically. During the task, 48 household objects (24 targets and 24 lures) were presented on a screen with a grey background for four seconds each. Participants indicated whether each object was 'old' or 'new' by providing a confidence rating for remembering or not remembering the object using a 6-point Likert scale (ranging from '1) very sure not seen' to '6) very sure seen'). For objects rated as 'old' (i.e. '4) a little sure seen', '5) quite sure seen', '6) very sure seen'), a room recall (RR) using a forced choice question followed. Participants selected the room where they remembered seeing the object from the following options: kitchen, bathroom, bedroom, or 'actually not seen.' If 'actually not seen' was chosen, the trial ended immediately. If a room was selected, participants provided a confidence rating for how sure or unsure they are to have seen the object in the specific room they selected by using a 6-point Likert scale (ranging from '1) very sure not seen' to '6) very sure seen') before the trial ended. The next object was then presented. No time limit was set for responses. The task took approximately 15 min to complete. Although confidence ratings for the memory of objects and memory of objects in a specific room were recorded, they were not analyzed in this study. Old/New ratings were utilized to assess general memory performance and to serve as a threshold for triggering the forced-choice question about the RR. This allowed us to use the total number of incongruent objects recognized by each participant when calculating the proportions of different memory systems to investigate semantic substitution in the end (see more details in section 2.5. *Statistics and data preparation*).

The primary focus of this study was on the RR data of incongruent objects, which served as a key indicator for distinguishing between underlying mechanisms during scenario construction, such as use of episodic memory (correct RR), semantic substitution (RR based on the related semantic category of the object), or potential random memory errors (unrelated RR). This approach enabled us to explore how participants reconstruct scenarios after encountering incongruent objects. For example, if a participant correctly recalled the toaster in the bathroom, this was classified as episodic memory. If the participant recalled the toaster in the kitchen (a semantically congruent location based on general semantic knowledge), this was classified as semantic memory. Conversely, if the toaster was recalled in an unrelated location, such as the bedroom, we assume it to be a potential random memory error. Thus, this was classified as wrong memory.

2.4. Cortisol administration, saliva sampling and analysis

In a double-blind, randomized design, half of the participants received two 10 mg tablets of cortisol (hydrocortisone; Hoechst) 30 min before retrieval (free recall specifically and 54 min before the recognition task) on day two, while the other half were given two visually identical placebos (Lichtenstein). This was done in accordance with our previous experiments (e.g., Hagedorn et al., 2021). We measured cortisol levels (nmol/l) at four different time points on day two: baseline, +30 min, +95 min, and + 105 min after tablet intake using Salivette devices (Sarstedt, Nümbrecht, Germany) along with momentary affect ratings (using the Positive Affect Negative Affect Scale, PANAS (Breyer & Bluemke, 2016)) recorded on a laptop. The saliva samples were stored at −20 °C until analyzed with a commercially available enzyme-linked immunosorbent assay (IBL International, Hamburg,

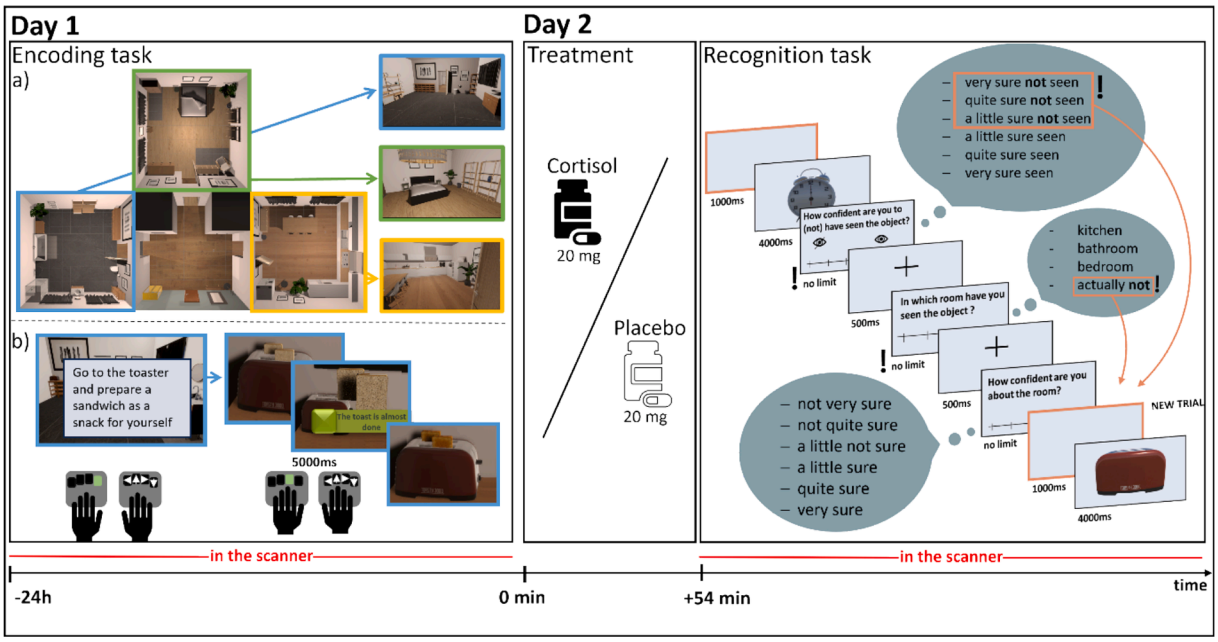


Fig. 1. Overview of the method. The left side of the figure shows the encoding task on day 1. a) For illustrative purposes, the virtual apartment is depicted from a bird's-eye view, with the three distinct rooms highlighted in yellow (kitchen), green (bedroom) and blue (bathroom). Arrows indicate the first-person perspectives participants experienced during encoding. b) Depiction of one exemplaric interaction with an object as a sequence. The images on the left show a representative view of the task instruction, and the images on the right demonstrate an example task within the virtual environment. The left hand was used for task display (index finger) and execution (middle finger), and the right hand fingers controlled navigation. For visualization purposes, the button pads are visualized below with different buttons required for the task highlighted in green or with arrows. After pressing the button, the object was elevated to eye level and a loading bar indicated the task progress (fixed duration of five seconds), after which the object transformed into its final state (e.g. a toaster with toasted bread). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
The right side of the figure shows the relevant components of day 2. The treatment box shows the randomized administration of 20 mg of a placebo or hydrocortisone. The recognition box, shows the sequence of events in a typical recognition task trial. Each trial began with a blank screen (highlighted in a bold frame) presentation for 1000 ms, followed by a 2D picture presentation (4000 ms) of a target or lure object. Targets were 24 objects from the encoding task. A bold exclamation mark highlighted the two crucial questions that could trigger an immediate new trial. The first rating screen was shown to indicate if the object presented before was old or new, including a corresponding confidence rating (no time limit). If an objects was rated as 'new', a new trial (highlighted in a bold orange frame) began immediately (illustrated with orange arrows). If recognized as old, a forced choice question in which room the objects were remembered followed (no time limit). Selecting the 'actually not seen' option prompted a new trial (illustrated with orange arrows). Indicating a recalled room elicited a corresponding confidence rating, concluding the trial. Between questions, a fixation cross (500 ms) was shown. The presentation of pictures was completely randomized. Answers were given using button pads located under participants' hands. Red lines highlight the tasks conducted inside the scanner. A timeline below displays the start of each task in refereence to the pharmacological onset.

Germany) in our in-house biochemical laboratory. Inter- and intra-assay coefficients of variation were below 10 %.

2.5. Statistics and data preparation

We prepared behavioral data for further analyses with Python 3.8 implementation in *Spyder* (Raybaut, 2009; Van Rossum and Drake, 2009). All statistical analyses were performed using R (R Core Team, 2020) with version 4.3.1 (2023–06-16) in R Studio (RStudio Team, 2019). If normal distribution or homoscedasticity was not given as examined using the Shapiro-Wilk test, the data was log-transformed when possible, or nonparametric alternative tests were used. In case of multiple comparisons, Bonferroni correction was used. The significance level was set to $\alpha = 0.05$. For repeated-measures analyses of variance (ANOVA), Greenhouse-Geisser corrected p-values were reported if the assumption of sphericity was violated. Treatment (cortisol vs. placebo) was included as a between-subjects factor. We conducted separate ANOVA with the repeated measurement factor time (baseline, +30 min, + 95 min, +105 min) to analyze differences in cortisol concentrations and negative affect ratings (and positive affect ratings on an exploratory level) on day two.

To ensure that participants complied with the recognition task and did not randomly press the same button to expedite completion, we calculated each participant's sensitivity measure dee prime (d') by using the standardized difference between hit rate and false alarm rate, in line with the method outlined by Macmillan and Creelman (1996). We then conducted a *t*-test on d' scores between groups to examine the impact of cortisol on episodic memory (performance). For the subsequent analyses, we focused exclusively on the targets and excluded the lures. First, to determine whether there was a difference in memory performance between congruently and incongruently placed objects, we conducted a *t*-test to compare the two conditions. Additionally, to assess the influence of cortisol on memory performance for target objects and for incongruent and congruent target objects individually, we performed three separate t-tests between the groups.

For comparability reasons with our behavioral study (Zöllner et al., 2022) and to explore how the object's characteristics, such as congruency and interaction, as well as cortisol influence, predict the correct room recall we used a logistic linear mixed model analysis. This approach allowed us to account for both individual subject effects and object effects by including them as random factors in our model. We followed the procedure outlined by Sommet and Morselli (2017), estimating the significance of predictors by computing 95 % confidence intervals (CI) and interpreting odd's ratio (OR); i.e., if the value 1 is part of the 95 % confidence interval, there is no significant effect of this predictor. Specifically, we focused on the hit rate (correct identification of a target object as 'old') and the correct (episodic) room recall and compared it across the two object characteristics and cortisol. In

addition to our preregistration, we used ANOVA to check how the proportion of used memory (episodic, semantic, wrong) was naturally distributed to get a first overview of the nature of memory and to check if the cortisol manipulation was successful (influence on episodic memories) and how it influenced semantic and wrong memories.

Finally, to investigate the primary focus of this study, i.e., scenario construction, specifically to accurately distinguish correct episodic memories from semantic memories or wrong errors, we calculated the percentage (expressed as decimal values) of incongruent objects recalled in each room based on the total number of remembered incongruent objects. For each participant, we calculated the proportion of episodic, semantic, and wrong memories by dividing the number of objects recalled in each category by the total number of incongruent objects recognized as 'old'. For example, if a participant recognized 10 out of 12 incongruent objects, with 6 remembered episodically, 3 semantically, and 1 wrongly, the resulting proportions would be 6/10 = 0.6 (episodically), 3/10 = 0.3 (semantically) and 1/10 = 0.1 (wrongly). Semantic substitution was considered valid only if the proportion of semantic memories was significantly higher than that of wrong memories, thus we conducted a *t*-test between proportion of semantic and proportion of wrong memories. This approach enable us to draw conclusions about scenario construction based on the room recall data while accounting for individual differences in general memory performance but maintaining randomization during encoding. Lastly, to test our central hypothesis regarding semantic substitution under elevated cortisol levels, we conducted a *t*-test including the difference of proportion of semantic and wrong memories between groups and a paired *t*-test to compare the proportion of semantic and wrong memories within each group. For comparability reasons with our behavioral study (Zöllner et al., 2022), we assessed an additional measurement for semantic substitution, the semantic bias, which is the quotient of the proportion of incongruent objects that are remembered episodically (INCONGepi) and the sum of episodically (INCONGepi) and semantically (INCONGsem) remembered incongruent objects $(\frac{INCONGepi}{INCONGepi+INCONGsem})$.

2.6. fMRI data acquisition and analyses

Functional and structural brain scans were acquired using a whole-body 3 T scanner (Philips Achieva 3.0 T X-Series, Philips, the Netherlands) with a 32-channel SENSE head coil. The structural images were obtained in a T1-weighted sequence at 1 mm isotropic resolution (field of view: 240 x 240 mm^2; slice thickness = 1 mm; voxel size 1 x 1 x 1 mm^3) with 220 transversally oriented slices covering the whole brain and a TA of 6 min 2 s. Blood oxygenation level-dependent (BOLD) contrast images were obtained with a T2*-weighted gradient echoplanar imaging EPI sequence (TR = 2.5 s; TE = 30 ms; flip angle: = 90°; field of view = 96 x 96 mm^2, slice thickness = 1 mm; 45 transversal slices in ascending order without slice gap; voxel size = 1 x 1 x 1 mm^3). Five dummy scans preceded each functional scan session to reach stable magnetization.

For preprocessing and statistical analyses, we used the software package Statistical Parametric Mapping (SPM12, Wellcome Department of Cognitive Neurology, London, UK) implemented in MATLAB 2021a (The MathWorks Inc., 2021). Preprocessing followed standard procedures, including realignment, slice time correction, co-registration of functional data to each participant's structural image, normalization to the standard space of the Montreal Neurological Institute (MNI) brain, and spatial smoothing using an 8-mm FWHM Gaussian kernel. For all statistical analyses, we used exploratory whole brain as well as region of interest (ROI) analyses, including brain regions identified as relevant for episodic memory (HPC, precuneus), semantic knowledge (angular gyrus, vmPFC, aPHC) and as conflict monitor for semantic substitution (ACC). The required masks were maximum probability masks with the probability threshold set to 0.25, taken from the Harvard-Oxford Cortical and Subcortical Structural Atlases provided by the Harvard Center for Morphometric Analysis (https://www.cma.mgh.harvard.edu/fsl_atlas.html) except for the vmPFC mask which consisted of a 5 mm sphere surrounding the peak voxel for schema-dependent connectivity between HPC and vmPFC during memory encoding (MNI coordinates x = -4, y = 24, z = -21, as indicated in a review of episodic and semantic memories (van Kesteren et al., 2010)). For the exploratory whole brain and ROI analyses, the significance threshold was set to $p \leq 0.05$ on voxel-level corrected for multiple testing (family-wise error (FWE) correction). ROI analyses were conducted using the small volume correction option of SPM12 with an initial intensity threshold of $p \leq 0.05$.

At the first level, we entered the following regressors: For congruent target (*CONG*) and distractor (*DIS*) stimuli, we specified three regressors each: semantically remembered (*CONGsem* and *DISsem*), wrongly remembered (*CONGwrong* and *DISwrong*) and not remembered (*CONGnot* and *DISnot*). The term 'not remembered' is used differently depending on the stimulus; it is equivalent to 'correct rejection' for distractor stimuli, while it refers to 'forgotten' for congruent stimuli. For incongruent target stimuli, we used four regressors: Correctly remembered episodically (*INCONGepi*), semantically substituted (*INCONGsem*), remembered wrongly (*INCONGwrong*), and not remembered (*INCONGnot*). We treated INCONGsem, CONGsem, and DISsem as separate regressors, although all of them refer to the recall of objects based on their semantic category because, for INCONG and CONG, objects differed in their initial encoding (congruently or incongruently), which could lead to differences in the retrieval process. Similarly, distractor objects, which were not encoded previously, may reflect a different retrieval process as well – potentially the closest to a pure semantic retrieval. For the three question screens, we differentiated between old/new rating screens (*ONRS*), room recall screens (*RRS*), and confidence ratings of room recall screens (*CRRS*). Button presses were also included as one regressor (*ALLRESPONSES*). The six realignment parameters were entered as covariates, and a high pass filter of 128 s was applied.

All parameters in the general linear model were modeled using a stick function convolved with the hemodynamic response function. The following first-level contrasts were created to test our hypotheses on the second level: (INCONGsem vs. INCONGwrong) and (INCONGepi vs. INCONGsem), additionally we analyzed the contrast (INCONGepi vs. INCONGwrong) in an exploratory manner. If participants displayed no episodic, semantic, and or wrong recollection of incongruent objects, they were excluded from the respective second-level contrast. Please note that we further included all participants with at least one valid trial (more information on the number of trials per memory proportion can be found in Table 3; results of additional post hoc analysis of participant with at least two valid trials can be found in the respective result section 3.7. *Neural responses* in the footnotes). On the second level, two-sample t-tests examined differences between the cortisol and the placebo group. In exploratory analyses, we entered significantly activated ROIs during semantic substitution as seed regions in functional connectivity analyses realized with psychophysiological interaction (PPI) analyses.

3. Results

3.1. Sample description

We tested 60 participants (30 for each of the two groups) aged between 18 and 34 years. No significant differences regarding age or body mass index (BMI) were found between participants in the placebo (mean age ± SD: 24.23 ± 4.03 years; BMI: 24.31 ± 2.85 kg/m^2) compared with the cortisol group (mean age ± SD: 24.53 ± 4.23 years; BMI: 24.00 ± 2.62 kg/m^2; both $p > 0.05$).

3.2. Salivary cortisol and affect

We excluded three participants from hormonal analyses due to displaying extremely high cortisol levels (larger than 1000 nmol/l) 30 min after cortisol intake, which most likely reflects some drug residue of the uncoated tablet in the participants' mouth. On day two, the ANOVA of salivary cortisol concentrations revealed a significant main effect of time ($F_{(3,165)} = 67.353$, $p < 0.001$, $\eta_p^2 = 0.064$), treatment ($F_{(1,55)} = 212.954$, $p < 0.001$, $\eta_p^2 = 0.004$), and a time x treatment interaction ($F_{(3,165)} = 99.578$, $p < 0.001$, $\eta_p^2 = 0.460$). Whereas groups did not differ at baseline ($p > 0.05$), cortisol was elevated 30, 95, and 105 min after cortisol compared to placebo administration (all $p < 0.001$; Table 1).

We acquired negative affect ratings parallel to cortisol measurements using the PANAS. The ANOVA revealed neither a significant main effect of treatment ($F_{(1,58)} = 2.15$, $p = 0.148$, $\eta_p^2 = 0.016$), nor a time x treatment interaction ($F_{(3,174)} = 1.09$, $p = 0.353$, $\eta_p^2 = 0.005$), but a significant main effect of time ($F_{(3,174)} = 3.41$, $p < 0.05$, $\eta_p^2 = 0.004$). Post hoc comparisons revealed an increase in negative affect from baseline to 95 min after tablet intake (($V = 928.5$, $p_{Tukey} < 0.05$), see Table 1). For reasons of completion, we exploratorily analyzed the positive affect in the same manner. ANOVA revealed neither a significant main effect of treatment ($F_{(1,58)} = 0.10$, $p = 0.749$, $\eta_p^2 = 0.0002$), nor a time x treatment interaction ($F_{(3,174)} = 0.71$, $p = 0.551$, $\eta_p^2 = 0.002$), but a significant main effect of time ($F_{(3,174)} = 24.39$, $p < 0.001$, $\eta_p^2 = 0.039$). Post hoc comparisons revealed a decrease in positive affect from baseline to 95 min ($V = 1512.5$, $p_{Tukey} < 0.001$) and to 105 min ($V = 1338.5$, $p_{Tukey} < 0.001$) after tablet intake. Taken together, all participants experienced a decrease in affect from the beginning to the end of the experiment. Additionally, cortisol administration successfully elevated cortisol concentrations but did not alter affect.

3.3. Cortisol effect on recognition (d') and memory performance for targets (congruent and incongruent)

Comparing the sensitivity measure d', i.e., hits – false alarms of target and lures, the *t*-test showed no statistical differences between the placebo and the cortisol group ($t(58) = 0.752$, $p = 0.455$, $d = 0.195$). Thus, cortisol did not impair general recognition performance.

Focussing only on the hit rate of targets, we found a similar pattern. During the recognition task, participants in both groups recognized on average approximately 18 (17.77 ± 3.14; $M \pm SD$) out of 24 objects. Conducting a *t*-test between groups, we found no significant cortisol influence on the number of recognized objects ($t(58) = 0.489$, $p = 0.626$, $d = 0.127$). Similarly, investigating the performance of memory for incongruent and congruent objects, groups did not differ, as revealed by using separate t-tests (both $p > 0.05$; see Table 2). Thus, cortisol did not influence the memory of congruent or incongruent objects specifically.

Table 1
(A) Mean (±SD) salivary cortisol concentrations at baseline and after 30 min, 95 min, and 105 min after cortisol administration (20 mg) or placebo. (B) Mean (±SD) of negative affect for the same timepoint as stated above. Data is shown separately for the placebo and cortisol group, respectively. The statistics are described in detail in the text.

	placebo ($n = 30$)	cortisol ($n = 27$)
(A) salivary cortisol (nmol/l)		
baseline	4.35 ± 3.87	3.35 ± 2.06
30 min after treatment	3.53 ± 2.09	119.27 ± 140.73
95 min after treatment	2.14 ± 1.17	31.92 ± 17.62
105 min after treatment	2.27 ± 1.21	25.08 ± 13.85
(B) negative affect		
baseline	1.24 ± 0.29	1.40 ± 0.38
30 min after treatment	1.20 ± 0.23	1.33 ± 0.38
95 min after treatment	1.18 ± 0.24	1.25 ± 0.31
105 min after treatment	1.24 ± 0.29	1.29 ± 0.33

Table 2
Mean (±SD) memory (hits and misses) of all, incongruent and congruent objects, shown for the placebo and cortisol group separately.

	all objects		congruent objects		incongruent objects	
	misses	hits	misses	hits	misses	hits
total	6.23 ± 3.14	17.77 ± 3.14	3.42 ± 2.03	8.58 ± 2.03	2.82 ± 1.64	9.18 ± 1.64
placebo	6.03 ± 3.37	17.97 ± 3.37	3.33 ± 2.28	8.67 ± 2.28	2.70 ± 1.62	9.30 ± 1.62
cortisol	6.43 ± 2.94	17.57 ± 2.94	3.50 ± 1.78	8.50 ± 1.78	2.93 ± 1.68	9.07 ± 1.68

3.4. Prediction of correct room recall by object characteristics and cortisol

To better understand the influence of object characteristics on memory and for comparability reasons with our behavioral study (Zöllner et al., 2022), we analyzed how factors such as interaction and congruency, along with the factor treatment, predicted accurate episodic memory reflected in the correct room recall using a multilevel model. Treatment did not predict correct room recall ($OR = 0.73$, 95 % $CI = [0.35, 1.55]$, $p = 0.419$). For congruency, we found that congruent objects had a higher probability of predicting the correct room recall compared to incongruent objects ($OR = 0.25$, 95 %$CI = [0.13, 0.47]$, $p < 0.001$). Although we did not find a significant effect on whether an object was an interaction object ($OR = 0.55$, 95 %$CI = [0.24, 1.26]$, $p = 0.158$), we found a significant interaction between object congruency and interaction of objects: congruent interaction objects were more likely to be correctly remembered to the correct room than incongruent non-interaction objects ($OR = 0.31$, 95 %$CI = [0.12, 0.80]$, $p < 0.05$). All other interaction effects were not significant ($p > 0.05$).

3.5. Memory of incongruent objects and cortisol

Only incongruent objects allowed us to assume the involvement of different memory systems (episodic vs. semantic) or memory errors (wrong). For each participant, we analyzed the percentage (indicated by decimal numbers in the following) of incongruent objects that were remembered in each room based on the total number of remembered incongruent objects (see Table 3 for individual data). We conducted a mixed ANOVA to understand better the overall distribution of memory errors across different forms of knowledge, not limited to semantic substitution. This was done in addition to our preregistered analysis to replicate the findings of our behavioral study (Zöllner et al., 2022) and as a confirmation of the results. The mixed ANOVA with the within-subjects factor memory proportion and the between-subjects factor treatment revealed a main effect for memory proportion ($F(2,116) = 60.781$, $p < 0.001$, $\eta_p^2 = 0.300$). Post hoc pairwise comparisons revealed a significantly higher proportion of episodic memories (0.55,7 ± 0.18; $M \pm SD$) compared to semantic memories (0.26,4 ± 0.17; $M \pm SD$), $t(174) = 9.783$, $p_{Tukey} < 0.001$) and wrong memories ((0.17,9 ± 0.12; $M \pm SD$), $t(174) = 12.873$, $p_{Tukey} < 0.001$). Besides, a higher proportion of semantic memories compared to wrong memories occurred ($t(174) = 3.090$, $p_{Tukey} < 0.005$). However, no treatment x memory proportion interaction was found ($F(2,116) = 1.379$, $p = 0.256$, $\eta_p^2 = 0.023$). All in all, participants displayed superior episodic recognition, instead of semantic or wrong recognition of incongruent objects. Thus, we can confidently affirm the proper execution of the modified recognition task and further validated our newly developed paradigm. In addition, we replicated the semantic substitution finding aligning with our previous research (Zöllner et al., 2022), whenever we observed no group differences using this approach.

Table 3
Overview of the total number of congruent (CONG) and incongruent (INCONG) objects recognized as old and the number of how often those incongruent objects were remembered episodically (epi), semantically (sem), or wrongly for each participant in the placebo (n = 30) and cortisol (n = 30) group separately.

placebo					cortisol				
CONG	INCONG				CONG	INCONG			
old	**old**	epi	sem	wrong	**old**	**old**	epi	sem	wrong
9	12	9	2	1	4	5	4	0	1
7	7	3	3	1	8	9	5	2	2
10	10	5	1	4	7	10	5	3	2
7	10	4	3	3	9	11	5	2	4
8	11	7	2	2	7	8	5	2	1
9	10	6	1	3	11	10	5	4	1
6	10	2	5	3	9	11	3	4	4
12	11	9	1	1	10	11	2	8	1
7	7	5	2	0	8	10	6	4	0
5	8	4	2	2	7	10	6	3	1
12	10	4	2	4	10	9	4	4	1
10	10	4	4	2	7	10	8	1	1
12	9	6	2	1	6	8	7	1	0
6	6	4	2	0	11	11	9	2	0
6	9	3	4	2	10	7	4	1	2
12	11	6	1	4	6	7	6	0	1
10	10	5	2	3	9	7	6	1	0
10	11	6	3	2	9	9	6	3	0
4	7	3	3	1	9	9	4	4	1
6	8	4	1	3	6	6	5	0	1
11	10	6	1	3	9	7	4	1	2
8	11	6	3	2	10	10	8	0	2
10	7	4	1	2	10	7	6	1	0
9	7	5	1	1	7	10	3	7	0
11	10	9	1	0	10	8	4	2	2
7	11	5	6	0	11	10	3	7	0
11	9	5	1	3	7	10	3	3	4
8	8	4	2	2	8	10	4	5	1
10	11	3	5	3	10	11	6	3	2
7	8	4	2	2	10	11	5	3	3

3.6. Semantic substitution and cortisol

Sticking to our preregistered method to specifically isolate the impact of cortisol on semantic substitution, defined as the ratio or difference of semantic to wrong memories (in case of episodically missing details), we first used the non-parametric Mann-Withney- *U* test, which however was not significant between groups ($p > 0.05$; Table 4). Secondly, we compared this proportion within groups using the Wilcoxon-signed rank test. Notably, the semantic substitution was present only in the cortisol group ($V = 243.5$, $p < 0.05$) but not in the placebo group ($V = 491.5$, $p_{Bonf} = 0.815$), hinting towards a cortisol-driven effect of semantic substitution (Fig. 2).

Additionally, for better comparability with our previous study, we compared the semantic bias (quotient of the proportion of incongruent objects that are episodically remembered and the sum of the proportion of episodic and semantically remembered objects), reflecting a favoring of semantic memories compared to wrong memories, between groups. The Mann-Whitney-U-Test could not provide a significant cortisol effect on the semantic bias ($W = 484$, $p = 0.696$). Participants in the placebo group showed a slightly more negligible semantic bias ($M = 1.07$, $SD = 0.8$) compared to the cortisol group ($M = 1.4$, $SD = 0.8$).

Table 4
Mean (±SD) percentage for proportion of episodic, semantic, or wrong memories of incongruent objects are shown for the placebo and cortisol group separately.

	proportion of memory type		
	episodic	semantic	wrong
total	55.7 ± 0.18	26.4 ± 0.17	17.9 ± 0.12
placebo	53.8 ± 0.15	25.0 ± 0.14	21.2 ± 0.12
cortisol	57.6 ± 0.20	27.8 ± 0.20	14.6 ± 0.12

3.7. Neural responses

Four participants had no semantic recollection (INCONGsem) and twelve had no wrong recollection (INCONGwrong), so we excluded these trials from second-level contrasts involving semantic or wrong memories. There was no missing data for episodic memories. For the comparison of episodic and semantic memories (INCONGepi vs. INCONGsem), the sample sizes were $n = 26$ for cortisol and $n = 30$ for placebo. For the comparison between episodic and wrong memories (INCONGepi > INCONGwrong), the sample sizes were $n = 22$ for cortisol and $n = 26$ for placebo. For semantic versus wrong memories (INCONGsem > INCONGwrong), sample sizes were $n = 18$ for cortisol and $n = 26$ for placebo. Additional post hoc analysis of the latter regressor (INCONGsem > INCONGwrong) was conducted, including only participants with at least two trials, resulting in sample sizes of $n = 8$ for cortisol and $n = 13$ for placebo.

3.7.1. Memory of incongruent objects and cortisol

None of the contrasts comparing episodic and semantic memories (INCONGepi > INCONGsem) or episodic with wrong memories (INCONGepi > INCONGwrong) showed any significant differences in activation at the whole brain level or within the hypothesized ROIs.

3.7.2. Semantic substitution and cortisol

For the semantic substitution contrast (*INCONGsem > INCONGwrong),* the whole brain analysis revealed a stronger differential activation in the gyrus rectus ($x = 2$, $y = 4$, $z = -18$). Additionally, as expected, we found increased activation in one of our ROIs, i.e. in the

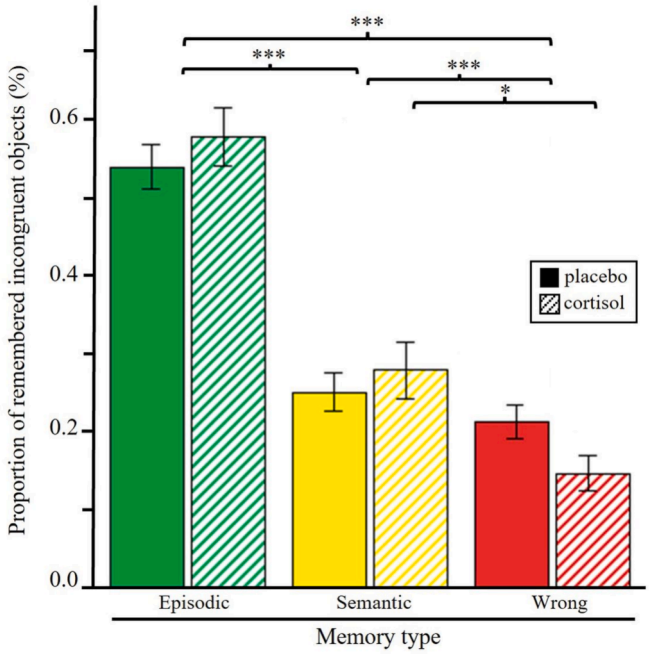


Fig. 2. The graph illustrates the percentage of the proportion of remembered incongruent objects (decimal format) for three different memory types: episodic, semantic, and wrong. The classification of episodic, semantic, or wrong memory is determined in which room participants remembered incongruent objects. Proportions are calculated by dividing the number of recalled objects for each classified memory type by the total number of recognized incongruent objects. The remembered incongruent objects are presented separately for the cortisol and placebo groups. In both groups, incongruent objects were most frequently remembered episodically rather than semantically or wrong. In the case of episodic failure, incongruent objects were more often remembered semantically than wrongly (semantic substitution). However, when analyzed separately by group, the semantic substitution was only evident in the cortisol group. Significance levels are denoted as * $pTukey < 0.05$ and *** $pTukey < 0.001$.

bilateral aPHG (Table 5).[2] Contrary to our hypothesis, the analysis of the contrast INCONGsem minus INCONGwrong indicated increased activation in the left HPC for the cortisol compared to the placebo group (see Fig. 3). No further activation in the other hypothesized ROIs was found. Exploratory PPI did not reveal any enhancing or decreasing functional connectivity between the left HPC and other brain regions.

4. Discussion

The study aimed to investigate the neural mechanisms underlying semantic substitution during the reconstruction of a past scenario and their modulation by the stress hormone cortisol. Our cortisol manipulation was successful, as evidenced by increased salivary cortisol levels in the cortisol group. Although we found no cortisol-related impairments on episodic memory (measured in d' and in the correct room recall of incongruent objects specifically) between groups, participants often failed to remember all details of the episode (household objects and where to find them) accurately. When those details were not remembered episodically, participants recalled them more frequently semantically than wrongly. Neural correlates during semantic substitution included the gyrus rectus and the aPHG. We found no differences in semantic substitution between groups. However, when analyzing groups separately, only participants in the cortisol group showed semantic substitution. At the neural level, we found increased activation in the left HPC during semantic substitution under cortisol, suggesting a cortisol-driven effect on semantic substitution. The semantic bias (quotient of the proportion of incongruent objects that are episodically remembered and the difference between episodically and semantically remembered objects) based on an approach of Zöllner et al. (2022) did not reveal a cortisol effect on semantic substitution between groups. On the subject level, we identified potential factors during encoding that may predict the successfulness of later episodic memory retrieval. While the factor cortisol did not predict episodic memory retrieval, congruent objects and congruent objects that had been interacted with, were the best predictors of correct episodic memory (correct room recall of objects) 24 h later.

Table 5
Peak-voxel statistics and localization for the contrasts of conditions including only incongruent objects (INCONG): INCONGsem vs INCONGwrong for (a) semantic substitution, the contrast INCONGepi vs INCONGsem for (b) correct episodic memory vs semantic knowledge and the contrast INCONGepivs INCONGwrong for (c) correct episodic memory vs false memories. The direction of the contrasts is marked. Cortisol effects and direction are reported. If a direction or cortisol effect is not listed, it was not significant.

Contrast	Brain structure	Cluster size	*x*	*y*	*z*	T_{max}	p_{corr}
(a) Semantic substitution							
INCONGsem > INCONGwrong	gyrus rectus (WB)	4	2	4	−18	5.80	0.017
	L aPHG	106	−14	−2	−22	4.41	0.005
	R aPHG	312	14	0	−22	4.40	0.005
Cortisol effects for INCONGsem > INCONGwrong (cortisol > placebo)	L hippocampus	221	−24	−36	−2	4.01	0.016
(b) correct episodic vs semantic knowledge INCONGepi > INCONGsem	No significant effect						
(c) correct episodic vs false memories INCONGepi > INCONGwrong	No significant effect						

Abbreviations: INCONGepi, incongruent objects remembered episodically; INCONGwrong, incongruent objects remembered wrongly; INCONGsem, incongruent objects remembered semantically; aPHG, anterior parahippocampal gyrus; L, left; R, right. The significance threshold was set to $p < 0.05$ (family-wise error-corrected for small volume correction and whole brain (WB) correction). Unless otherwise indicated, the regions shown are regions of interest (ROIs).The peak voxel from the WB analysis was labeled based on the Harvard-Oxford Cortical and Subcortical Structural Atlas. All coordinates (*x, y, z*) are given in MNI space.

[2] Further analyis including only participants with at least two trials, revealed no significant results for the whole brain and ROI analyses.

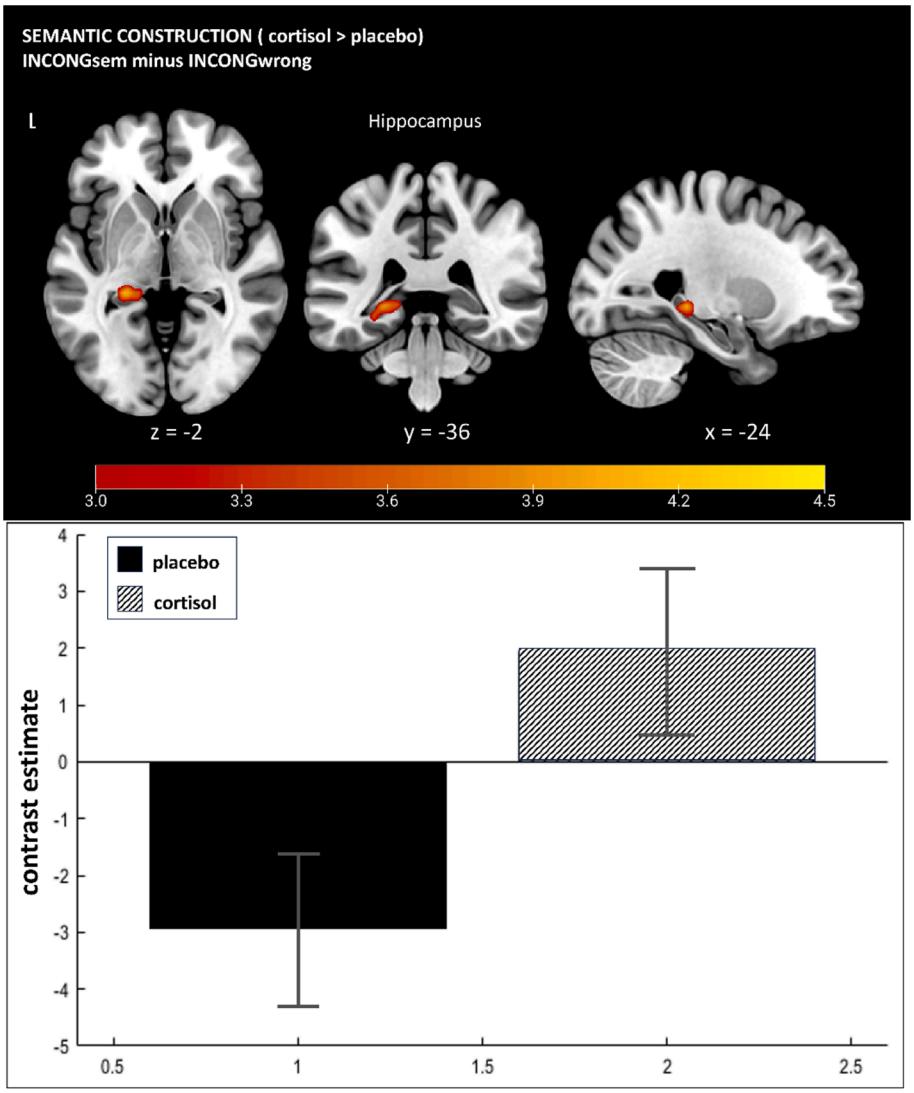


Fig. 3. Neural responding for the contrast semantically remembered incongruent objects (INCONGsem) minus wrongly remembered incongruent objects (INCONGwrong) during the recognition task. The depicted slice was selected according to the peak voxel of the left hippocampus. Data are presented on the standard MNI brain template and thresholded to T > 3.0 (see color bar for exact T values). The plot represents contrast estimates of the peak voxel. Abbreviations: L, left. During trials in the recognition task, in which incongruent objects were remembered semantically compared to wrongly, cortisol increased activation of the left hippocampus relative to placebo.

4.1. Cortisol effect on recognition (d') and memory performance for targets (congruent and incongruent)

In contrast to our finding of no impairment on recognition memory (d') after cortisol administration, according to previous studies, high concentrations of stress hormones before retrieving information can negatively affect memory retrieval (Het et al., 2005; Wolf, 2017). However, a review pointed out that recognition tasks seem less vulnerable to stress-induced impairments than free recall and cued recall tasks (Gagnon & Wagner, 2016), which might explain why we found no impairing cortisol effect on general memory performance in the recognition task. In addition, we used neutral stimuli only, while previous studies suggested that stress hormones have a smaller impact on the retrieval of neutral stimuli compared to emotional stimuli (Kuhlmann & Wolf, 2006; Shields et al., 2017; Wolf, 2009).

4.2. Prediction of correct room recall by object characteristics and cortisol

Our findings show that congruent objects had a higher probability of predicting the correct room recall compared to incongruent objects. This finding of increased memory for congruent objects is consistent with the congruency effect, which suggests that people tend to remember congruent or schematic information better compared to non-schematic or incongruent information (van Kesteren et al., 2012; Webb et al., 2016; Zöllner et al., 2022). The current results also replicate results of our behavioral study (Zöllner et al., 2022), which revealed that participants were more likely to sort objects to the correct room when objects were congruent to prior knowledge and relevant to the task at hand simultaneously. Besides, a possible explanation for why participants remembered objects that are relevant to the task at hand and schema congruent simultaneously might be due to a combination of the congruency effect and a bottom-up bias, which facilitates the retrieval of goal-relevant knowledge.

4.3. Semantic substitution (and neural response)

Our study revealed that the gyrus rectus, a subregion of the orbitofrontal cortex (OFC), was involved during usage of semantic knowledge in memory retrieval. This is in line with signs of impaired semantic memory after resection of the gyrus rectus (Szatkowska et al., 2004).

The association of the gyrus rectus with semantic knowledge might be relevant due to its role in response inhibition (Szatkowska et al., 2007) and suppression of irrelevant memories (Schnider, 2003). Thus, it is likely that the increased activation of the gyrus rectus helps to inhibit the accidental choice of the wrong room (which is chosen the least amount of times), which is unrelated to the object in any way.

Furthermore, we found increased activation in the aPHG for semantic compared to wrong object location recognition, which is in line with its association with familiarity (Eichenbaum et al., 2007). For example, more involvement of the anterior HPC during retrieving picture pairs was found in a schema-consistent compared to a schema-inconsistent condition (Guo & Yang, 2020). Beyond its role in memory retrieval, the PHG has mainly been examined in terms of spatial navigation, thus for encoding spatial scenes (Köhler et al., 2002), as well as during retrieval of spatial, compared to nonspatial, episodic information (Hayes et al., 2004, 2007). More specifically, the PHG is involved in the retrieval of spatial representations from semantic knowledge (Hoscheidt et al., 2010; Ryan et al., 2010). This aligns with the idea that when constructing a scenario, semantic information is used to supplement missing details, which can be incorporated into memory. Altogether, these findings add to our understanding of how the brain processes memories, revealing that it retrieves memories in a constructive and intertwined manner during scenario construction (Cheng et al., 2016).

4.4. Semantic substitution under cortisol (and neural response)

The results showed no differences in semantic substitution between groups at a behavioral level. However, when we analyzed the groups separately, we found that only participants in the cortisol group exhibited semantic substitution. Additionally, we observed increased activation in the left HPC during semantic memory retrieval in the cortisol group. This finding contradicts previous research suggesting HPC activation is reduced during episodic memory retrieval after cortisol administration (Oei et al., 2007). However, we need to address that typically reduced HPC activation after cortisol administration is associated with impaired episodic memory retrieval (Wolf, 2017), but we did not find this effect in our study. Thus, under the influence of the stress hormone cortisol, semantic knowledge might be used during an even earlier stage in the reconstructive process to compensate for missing details, instead of in a sequential manner (i.e., only after a completely disrupted episodic memory trace).

According to the SCM (Cheng et al., 2016) semantic knowledge becomes relevant when the gist is successfully retrieved but some episodic details are not available. Thus, one explanation for increased HPC activation during semantic compared to wrong sorting may indicate successful retrieval of the gist at least. Perhaps HPC activation is a sign of an unsuccessful search in episodic memory under the influence of cortisol, suggesting a compensatory effect of the HPC during scenario construction. Similar to our findings, stress led to increased HPC activation during the processing and memory of schema-related compared to schema-unrelated words (Vogel et al., 2018). Thus, increased HPC activation during memory retrieval in times of high cortisol levels might be more specific to semantic memories, whereas cortisol reduces HPC involvement during episodic memory retrieval. Please note that the post hoc analysis of imaging data on semantic substitution, including only participants who completed at least two trials for the regressor (INCONGsem > INCONGwrong), showed no significant effects on the neural level. This could be due to power issues of the small sample size of 21 participants (cortisol: $n = 8$, placebo: $n = 13$). This finding highlights the need for more trials to account for stable group sizes and improve statistical power for detecting semantic substitution at the neural level. Thus, caution is needed in interpreting our preliminary data of neural correlates.

4.5. Limitations

We only included men in our study; thus, the results cannot be readily generalized to women. Cortisol has been found to impair memory retrieval in naturally cycling women but not in women taking hormonal contraceptives (Jentsch et al., 2022). Therefore, further studies should include women comparing those under hormonal treatment and without to further investigate scenario construction in the general population. Importantly, semantic substitution (predominance of semantic vs wrong, in case of incomplete episodic memory), was present in the cortisol group only and not in the placebo group, but no group difference occurred directly. It is likely that, indeed, there is a cortisol-driven effect on semantic substitution but not large enough that it achieved significance between groups, which might be a power issue. No cortisol effect was found when the semantic bias was calculated. It is crucial to select the most suitable approach to measure semantic substitution. Future research incorporating comparable complex designs as ours is needed to investigate the intertwined relation between episodic and semantic memories engaged during a scenario, using appropriate calculation methods. Please note that even after excluding participants with no recollection of semantic or wrong memories, we still had some participants with only one trial in the respective regressors. Further analysis of imaging data, which included participants with at least two trials, revealed no significant activation in the whole brain or ROI analyses. This raises concerns about reliability and statistical power, especially in the imaging data, thereby indicating that our results should be interpreted as preliminary in nature. Future studies should include a higher number of trials to enhance the potential for differentiation between semantic substitution and unrelated memory errors, thus improving the underlying statistical power. Additionally, it is advisable to use a more conservative statistical threshold than $p \leq 0.05$ (FWE-corrected). Since we found no impairing cortisol effect on memory of recognition (measured in d') future research might profit from rethinking additions to a recognition test for investigating scenario construction with regard to the influence of cortisol on episodic memory.

5. Conclusion

Our results contribute to the view that even if episodic and semantic memory systems are structurally independent, they are functionally related in some parts, which advocates for an intertwined episodic-semantic constructive view on their involvement in memory retrieval, in men only. Even though imaging results should be interpreted with caution, we found first hints that neural correlates during scenario construction encompassed two brain regions, associated with involvement in semantic knowledge, i.e. the aPHG (familiarity) and the gyrus rectus, possibly reflecting involvement of semantic knowledge via inhibition of the accidental choice of the false room responses. Furthermore, after cortisol administration, missing details are more often substituted with semantic knowledge than unrelated information (guessing), when only the gist of an episode was remembered. Under the influence of cortisol, compensatory HPC activation during semantic compared to wrong memories occurred, enabling the completion of episodically impoverished events, at least in men.

The study lays a strong foundation for future research on semantic substitution. It offers initial insight into the neural correlates of semantic substitution (under cortisol influence) and paves the way for further imaging studies as well as methodological implications of relevant factors in investigating scenario construction. We encourage researchers to include women and all genders in their future studies on semantic substitution. Additionally, this research could lead to further investigations into the effects of psychological stressors on semantic substitution.

CRediT authorship contribution statement

Nicole Klein: Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Carina Zöllner:** Methodology, Conceptualization. **Tobias Otto:** Resources, Methodology. **Oliver Tobias Wolf:** Resources, Methodology. **Christian Josef Merz:** Writing – review & editing, Validation, Supervision, Project administration, Methodology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We gratefully acknowledge the help of Max Wedler and Henry Soldan during data collection and recruitment of participants. Moreover, we thank PHILIPS Germany for their invaluable technical support. Funding for this work was supported by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) within the FOR 2812:'Constructing scenarios of the past' under grant no. 419039274 (OTW) and within the SFB 1280: Extinction Learning under grant no. 316803389 – SFB1280 (projects A09 to OTW and CJM and INF to TO). The DFG had no role in study design, collection, analysis, interpretation of data, manuscript writing, or submitting the manuscript for publication.

Data availability

Data will be made available on request.

References

Barry, D. N., & Maguire, E. A. (2019). Remote memory and the hippocampus: a constructive critique. *Trends in Cognitive Sciences, 23*(2), 128–142. https://doi.org/10.1016/j.tics.2018.11.005

Bein, O., Livneh, N., Reggev, N., Gilead, M., Goshen-Gottstein, Y., & Maril, A. (2015). Delineating the effect of semantic congruency on episodic memory: the role of integration and relatedness. *PLOS ONE, 10*(2), Article e0115624. https://doi.org/10.1371/journal.pone.0115624

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences, 15*(11), 527–536. https://doi.org/10.1016/j.tics.2011.10.001

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature, 402*, 179–181. https://doi.org/10.1038/46035

Breyer, B., & Bluemke, M. (2016). Deutsche Version der positive and negative affect schedule PANAS (GESIS Panel). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. https://doi.org/10.6102/ZIS242

Brod, G., & Shing, Y. L. (2019). A boon and a bane: Comparing the effects of prior knowledge on memory across the lifespan. *Developmental Psychology, 55*(6), 1326–1337. https://doi.org/10.1037/dev0000712

G. Brod M. Werkle-Bergner Y.L. Shing The influence of prior knowledge on memory: A developmental cognitive neuroscience perspective. Frontiers in Behavioral Neuroscience 2013 7.

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain, 129*(3), 564–583. https://doi.org/10.1093/brain/awl004

Cheng, S., Werning, M., & Suddendorf, T. (2016). Dissociating memory traces and scenario construction in mental time travel. *Neuroscience and Biobehavioral Reviews, 60*, 82–89. https://doi.org/10.1016/j.neubiorev.2015.11.011

de Kloet, E. R., Joëls, M., & Holsboer, F. (2005). Stress and the brain: From adaptation to disease. *Nature Reviews Neuroscience, 6*(6), 463–475. https://doi.org/10.1038/nrn1683

Dedovic, K., Duchesne, A., Andrews, J., Engert, V., & Pruessner, J. C. (2009). The brain and the stress axis: The neural correlates of cortisol regulation in response to stress. *NeuroImage, 47*(3), 864–871. https://doi.org/10.1016/j.neuroimage.2009.05.074

Devitt, A. L., Addis, D. R., & Schacter, D. L. (2017). Episodic and semantic content of memory and imagination: A multilevel analysis. *Memory & Cognition, 45*(7), 1078–1094. https://doi.org/10.3758/s13421-017-0716-1

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends in Cognitive Sciences, 11*(9), 379–386. https://doi.org/10.1016/j.tics.2007.08.001

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience, 30*. https://doi.org/10.1146/annurev.neuro.30.051606.094328

Gagnon, S. A., & Wagner, A. D. (2016). Acute stress and episodic memory retrieval: Neurobiological mechanisms and behavioral consequences. *Annals of the New York Academy of Sciences, 1369*(1), 55–75. https://doi.org/10.1111/nyas.12996

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia, 53*, 104–114. https://doi.org/10.1016/j.neuropsychologia.2013.11.010

Guo, D., & Yang, J. (2020). Interplay of the long axis of the hippocampus and ventromedial prefrontal cortex in schema-related memory retrieval. *Hippocampus, 30*(3), 263–277. https://doi.org/10.1002/hipo.23154

Hagedorn, B., Wolf, O. T., & Merz, C. J. (2021). Stimulus-based extinction generalization: neural correlates and modulation by cortisol. *International Journal of Neuropsychopharmacology, 24*(4), 354–365. https://doi.org/10.1093/ijnp/pyaa085

Hayes, S. M., Nadel, L., & Ryan, L. (2007). The effect of scene context on episodic object recognition: Parahippocampal cortex mediates memory encoding and retrieval success. *Hippocampus, 17*(9), 873–889. https://doi.org/10.1002/hipo.20319

Hayes, S. M., Ryan, L., Schnyer, D. M., & Nadel, L. (2004). An fMRI study of episodic memory: retrieval of object, spatial, and temporal information. *Behavioral Neuroscience, 118*(5), 885–896. https://doi.org/10.1037/0735-7044.118.5.885

Herman, J. P., Ostrander, M. M., Mueller, N. K., & Figueiredo, H. (2005). Limbic system mechanisms of stress regulation: Hypothalamo-pituitary-adrenocortical axis. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 29*(8), 1201–1213. https://doi.org/10.1016/j.pnpbp.2005.08.006

Het, S., Ramlow, G., & Wolf, O. (2005). A meta-analytic review of the effects of acute cortisol administration on human memory. *Psychoneuroendocrinology, 30*(8), 771–784. https://doi.org/10.1016/j.psyneuen.2005.03.005

Hoscheidt, S. M., Nadel, L., Payne, J., & Ryan, L. (2010). Hippocampal activation during retrieval of spatial context from episodic and semantic memory. *Behavioural Brain Research, 212*(2), 121–132. https://doi.org/10.1016/j.bbr.2010.04.010

Jentsch, V. L., Pötzl, L., Wolf, O. T., & Merz, C. J. (2022). Hormonal contraceptive usage influences stress hormone effects on cognition and emotion. *Frontiers in Neuroendocrinology, 67*, Article 101012. https://doi.org/10.1016/j.yfrne.2022.101012

Köhler, S., Crane, J., & Milner, B. (2002). Differential contributions of the parahippocampal place area and the anterior hippocampus to human memory for scenes. *Hippocampus, 12*(6), 718–723. https://doi.org/10.1002/hipo.10077

Kuhlmann, S., & Wolf, O. T. (2006). Arousal and cortisol interact in modulating memory consolidation in healthy young men. *Behavioral Neuroscience, 120*(1), 217–223. https://doi.org/10.1037/0735-7044.120.1.217

Lupien, S. J., Maheu, F., Tu, M., Fiocco, A., & Schramek, T. E. (2007). The effects of stress and stress hormones on human cognition: Implications for the field of brain and cognition. *Brain and Cognition, 65*(3), 209–237. https://doi.org/10.1016/j.bandc.2007.02.007

Macmillan, N. A., & Creelman, C. D. (1996). Triangles in ROC space: History and theory of "nonparametric" measures of sensitivity and response bias. *Psychonomic Bulletin & Review, 3*(2), 164–170. https://doi.org/10.3758/BF03212415

Merz, C. J., & Wolf, O. T. (2017). Sex differences in stress effects on emotional learning: sex differences, stress, and emotional learning. *Journal of Neuroscience Research, 95*(1–2), 93–105. https://doi.org/10.1002/jnr.23811

Moscovitch, M. (1994). Memory and working with memory: Evaluation of a component process model and comparisons with other models. *Memory Systems, 1994*(369–394), 224.

Oei, N. Y. L., Elzinga, B. M., Wolf, O. T., de Ruiter, M. B., Damoiseaux, J. S., Kuijer, J. P. A., … Rombouts, S. A. R. B. (2007). Glucocorticoids decrease hippocampal and prefrontal activation during declarative memory retrieval in young men. *Brain Imaging and Behavior, 1*(1–2), 31–41. https://doi.org/10.1007/s11682-007-9003-2

Preston, A. R., & Eichenbaum, H. (2013). Interplay of hippocampus and prefrontal cortex in memory. *Current Biology, 23*(17), R764–R773. https://doi.org/10.1016/j.cub.2013.05.041

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature, 435*(7045), Article 7045. https://doi.org/10.1038/nature03687

R Core Team. *R: A language and environment for statistical computing* (Version R version 4.3.1. (2023-06-16)) [Computer software]. R Foundation for Statistical Computing. http://www.R-project.org/.

Raybaut, P. (2009). *Spyder-documentation*. Available Online at: Pythonhosted. Org.

Robin, J., & Moscovitch, M. (2017). Details, gist and schema: Hippocampal–neocortical interactions underlying recent and remote episodic and spatial memory. *Current Opinion in Behavioral Sciences, 17*, 114–123. https://doi.org/10.1016/j.cobeha.2017.07.016

RStudio Team. (2019). *R Studio (Version R version 4.4.0* ((2024–04-24 ucrt))). PBC: [Computer software]. RStudio.

Rugg, M. D., & Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Current Opinion in Neurobiology, 23*(2), 255–260. https://doi.org/10.1016/j.conb.2012.11.005

Ryan, L., Lin, C., Ketcham, K., & Nadel, L. (2010). The role of medial temporal lobe in retrieving spatial and nonspatial relations from episodic and semantic memory. *Hippocampus, 20*(1), 11–18. https://doi.org/10.1002/hipo.20607

Schnider, A. (2003). Spontaneous confabulation and the adaptation of thought to ongoing reality. *Nature Reviews Neuroscience, 4*(8), 662–671. https://doi.org/10.1038/nrn1179

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments, 10*(3), 266–281. https://doi.org/10.1162/105474601300343603

Sekeres, M. J., Winocur, G., & Moscovitch, M. (2018). The hippocampus and related neocortical structures in memory transformation. *Neuroscience Letters, 680*, 39–53. https://doi.org/10.1016/j.neulet.2018.05.006

Shields, G. S., Sazma, M. A., McCullough, A. M., & Yonelinas, A. P. (2017). The effects of acute stress on episodic memory: A meta-analysis and integrative review. *Psychological Bulletin, 143*(6), 636–675. https://doi.org/10.1037/bul0000100

Sommet, N., & Morselli, D. (2017). Keep calm and learn multilevel logistic modeling: a simplified three-step procedure using Stata, R, Mplus, and SPSS. *International Review of Social Psychology, 30*(1), Article 1. https://doi.org/10.5334/irsp.90

Szatkowska, I., Szymańska, O., Bojarski, P., & Grabowska, A. (2007). Cognitive inhibition in patients with medial orbitofrontal damage. *Experimental Brain Research, 181*(1), 109–115. https://doi.org/10.1007/s00221-007-0906-3

Szatkowska, I., Szymańska, O., & Grabowska, A. (2004). The role of the human ventromedial prefrontal cortex in memory for contextual information. *Neuroscience Letters, 364*(2), 71–75. https://doi.org/10.1016/j.neulet.2004.03.084

van Kesteren, M. T. R., Fernández, G., Norris, D. G., & Hermans, E. J. (2010). Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. *Proceedings of the National Academy of Sciences, 107*(16), 7550–7555. https://doi.org/10.1073/pnas.0914892107

van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences, 35*(4), 211–219. https://doi.org/10.1016/j.tins.2012.02.001

Van Rossum, G., & Drake, F. L., Jr. (Eds.). (2009). *Python 3 Reference Manual.* Createspace: (Python Documentation Manual Part 2).

Vogel, S., Kluen, L. M., Fernández, G., & Schwabe, L. (2018). Stress leads to aberrant hippocampal involvement when processing schema-related information. *Learning & Memory, 25*(1), 21–30. https://doi.org/10.1101/lm.046003.117

Webb, C. E., Turney, I. C., & Dennis, N. A. (2016). What's the gist? The influence of schemas on the neural correlates underlying true and false memories. *Neuropsychologia, 93*, 61–75. https://doi.org/10.1016/j.neuropsychologia.2016.09.023

Wolf, O. T. (2009). Stress and memory in humans: Twelve years of progress? *Brain Research, 1293*, 142–154. https://doi.org/10.1016/j.brainres.2009.04.013

Wolf, O. T. (2017). Stress and memory retrieval: Mechanisms and consequences. *Current Opinion in Behavioral Sciences, 14*, 40–46. https://doi.org/10.1016/j.cobeha.2016.12.001

Zöllner, C., Klein, N., Cheng, S., Schubotz, R. I., Axmacher, N., & Wolf, O. T. (2022). Where was the toaster? A systematic investigation of semantic construction in a new virtual episodic memory paradigm. *Quarterly Journal of Experimental Psychology, 17470218221116610*. https://doi.org/10.1177/17470218221116610