Biological Psychology 209 (2026) 109345

Contents lists available at ScienceDirect

Biological Psychology

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

The influence of acute stress on scenario construction

Nicole Klein, Henry David Soldan, Oliver Tobias Wolf, Christian Josef Merz *

Department of Cognitive Psychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr University Bochum, Universitätsstr. 150, Bochum 44780, Germany

ARTICLE INFO

Keywords:
Cortisol
Episodic memory
Prior knowledge
Semantic memory
Stress hormones

ABSTRACT

Stress impairs episodic memory retrieval. During scenario construction, semantic knowledge can complete episodic memories. However, it remains unclear how acute stress influences the interplay between episodic and semantic memories. In this preregistered study, 88 participants (44 women) were exposed to 24 household objects within a computer-based virtual 3D apartment, half of which they interacted with. The objects were placed either congruently (a toothbrush in the bathroom) or incongruently with semantic knowledge (a toaster in the bathroom, allowing to test the semantic contribution). The following day, acute stress induction took place before recognition of objects and their respective rooms. The stress induction successfully elicited stress responses, as indicated by physiological markers and self-reports, including an increase in salivary cortisol concentrations. Incorrect episodic memories were more often substituted by semantically congruent than unrelated rooms, indicating semantic construction. However, stress or sex did not influence semantic construction. Additionally, confidence ratings for correct episodic responses were the highest overall, while semantic errors received lower confidence than unrelated errors. Interestingly, acute stress increased confidence in unrelated errors in women. Overall, these findings support a constructive account of episodic memory retrieval involving completion with semantic information. Acute stress does not seem to directly influence semantic construction in this task. Future research should consider comparing different stress responses over time and their impact on the interplay between episodic and semantic memories.

1. Introduction

Stress is a ubiquitous part of everyday life. Remembering where you placed your car keys during a hectic morning is crucial for future action planning and ensuring you arrive at work on time. However, memories can sometimes be inaccurate or unreliable, leading to disruptions in daily routines. For example, misremembering the location of your car keys and subsequently searching for them in the wrong places can cause delays. Importantly, these inaccurate memories may not occur by chance but could be the result of systematic underlying mechanisms during memory reconstruction. For instance, prior semantic knowledge - knowing that your bag is the typical location of your car keys - can shape our recollection of events.

Traditionally, episodic memory has been defined by the binding of "what, where, and when" information into a single integrated representation of a personally experienced event (Tulving, 1972, 2002). At the same time, the influence of semantic knowledge on episodic memory has been a central focus of memory research for decades, leading to theoretical models that describe their interdependence and to numerous behavioral studies (Bartlett, 1932; Cheng et al., 2016; Macias & Persaud, 2024). Recent theoretical advancements, such as the Scenario Construction Framework (Cheng et al., 2016; Cheng, 2024) propose a dynamic interplay. According to this view, episodic memory relies on traces containing only the minimal gist of an experience, rather than a high-fidelity record. During retrieval, these sparse episodic traces trigger the reconstruction of semantic representations, which are further enriched with general semantic knowledge to generate a coherent scenario of the past (Cheng, 2024). In the current paradigm, we investigate this interplay by focusing on the integration of objects ("what") into their corresponding rooms. Crucially, we conceptualize the "room" primarily as a semantic contextual category (e.g., a bathroom characterized by thematically related objects) rather than as fine-grained spatial "where" information (e.g., precise coordinate locations within the room). This design allows us to observe how semantic expectations – such as the knowledge that a toaster belongs in a kitchen – shape the reconstructive process when episodic gist is weakly encoded or temporarily inaccessible.

Most of the work on the interaction of episodic and semantic memory

* Corresponding author.
E-mail addresses: nicole.klein-g5y@rub.de (N. Klein), henry.soldan@rub.de (H.D. Soldan), oliver.t.wolf@rub.de (O.T. Wolf), christian.j.merz@rub.de (C.J. Merz).

https://doi.org/10.1016/j.biopsycho.2026.109345
Received 27 February 2026; Received in revised form 11 June 2026; Accepted 9 July 2026
Available online 10 July 2026

has used simplified paradigms, such as the presentation of pictures in semantically congruent or incongruent contexts (Küppers & Bayen, 2014; Lampinen et al., 2001; Webb et al., 2016) or word lists from specific semantic categories in the Roediger-McDermott paradigm, where participants often recall semantically related but non-presented words (Ballou & Sommers, 2008; Chang & Brainerd, 2021; Craik & Tulving, 1975; Roediger & McDermott, 1995). However, these studies typically focus on short retention intervals, ranging from minutes to hours, and rely on artificial stimuli such as image pairs or word lists. Less is known about how semantic and episodic memory interact after longer consolidation periods. Recent evidence suggests a reliance on semantic knowledge in case of episodic memory failure, reflected in so-called “semantic construction” errors during retrieval, using a video game-like encoding task with congruent and incongruent objects in a realistic setting (Zöllner et al., 2023). On the neural level, semantic construction was shown to be associated with anterior parahippocampal and gyrus rectus activation (Klein et al., 2025). Yet, it remains unclear how acute stress influences the interplay between semantic and episodic memory.

After exposure to a stressful situation, the rapid activation of the sympathetic nervous system (SNS) results in the release of catecholamines such as (nor)adrenaline from the adrenal medulla. In parallel, the slower-acting hypothalamus-pituitary-adrenocortical (HPA) axis gets activated, leading to the release of glucocorticoids (primarily cortisol in humans with peak levels occurring ~25 min after stress onset (Joëls & Baram, 2009)). Catecholamines bind to adrenergic receptors, whereas glucocorticoids bind to mineralocorticoid and glucocorticoid receptors (Hermans et al., 2014; Ulrich-Lai & Herman, 2009) located in subcortical structures such as the amygdala and hippocampus and in the prefrontal cortex. Thus, these stress hormones are able to modify brain activation related to memory processes (Schwabe et al., 2012; Wolf, 2009). Importantly, recent findings showed that administration of the stress hormone cortisol increased hippocampal activation during semantic construction (Klein et al., 2025). Given this first evidence and extensive research on the impairing impact of stress on episodic memory retrieval (Shields et al., 2017; Wolf, 2017), it could be expected that acute stress also influences scenario construction by disrupting the availability of episodic memories, resulting in enhanced reliance on semantic memories.

This preregistered behavioral study aimed to investigate the influence of stress induction on the balance between episodic memory and semantic information during the reconstruction of a past scenario. Therefore, we adapted a previously created computer-based virtual 3D apartment in which participants interacted with household objects that were placed in accordance with semantic knowledge (toothbrush in the bathroom) or at odds (toaster in the bathroom; Zöllner et al., 2023). One day later, we induced acute stress or a control condition prior to a modified recognition test, including a room recall meant to disentangle markers of episodic memory (gist) from specific memory errors (semantic or unrelated) during scenario reconstruction.

We assumed stress to increase cortisol concentrations and alpha-amylase levels (as markers for HPA axis and SNS activation) as well as heart rate (HR) and blood pressure (BP) and to influence ratings for both negative and positive affect. Acute stress should reduce episodic memory accuracy during retrieval, resulting in increased semantic construction. On an exploratory level, we were interested in the influence of object characteristics such as congruency and task-relevance (Zöllner et al., 2023) and sex differences on scenario construction, since a growing body of work suggests that sex hormones modify stress reactivity and related memory processes (Jentsch et al., 2022; Kudielka & Kirschbaum, 2005).

2. Material and methods

2.1. Participants

We determined the sample size using G*Power 3.1 (Faul et al., 2009), assuming a medium sized effect (d=0.5) of acute psychosocial stress on episodic memory retrieval, as reported in a similar study by Smeets (2011) for our preregistration (https://osf.io/d65vg/?view_only=f3ff592f27994690ad0416dc2eb974b8). To detect a main effect of stress on memory retrieval with a desired power of 1-β= .80 and a type I error of α= .05, we required 87 participants, which we increased to 88 to ensure equal sizes (44 per group, balanced by sex). While our primary power calculation was based on this main effect, we analyzed the data using a 2 × 2 design to exploratorily investigate the potential modifying role of sex, given evidence of sex differences in stress reactivity (Kudielka & Kirschbaum, 2005). We included healthy, non-smoking, right-handed participants aged 18–35 years without motion sickness sensitivity. Participants were instructed to avoid food, drinks (except water), and exercise two hours before sessions. To account for influence of hormonal contraceptives in interaction with stress hormones on memory (Jentsch et al., 2022), we included only women not using hormonal contraceptives and tested naturally cycling women exclusively outside their menses. Written informed consent was obtained prior to testing. All procedures adhered to the Declaration of Helsinki and were approved by the ethics committee of the Medical Faculty at Ruhr University Bochum (reg.-nr.: 18–6368).

2.2. Encoding task

For this experiment, we modified the episodic virtual environment (EVE) task, as previously implemented (Zöllner et al., 2023). In the EVE task, participants navigated a computer-based virtual 3D apartment in a first-person perspective using keyboard controls to complete twelve tasks with household objects. At the beginning of the encoding phase, participants were visually presented with a general "cover story" instruction: they were told that they needed to tidy up and prepare the virtual apartment for an upcoming guest, and their task was to await and carry out individual, object-specific actions. The apartment had three rooms - a kitchen, bathroom, and a bedroom - each equidistant from one another. A total of 24 objects were pseudorandomly placed in these rooms, with four objects placed congruently to each room’s purpose (e.g., a toothbrush in the bathroom) and four incongruently (e.g., a toaster in the bathroom). To manipulate task-relevance, only half of the 24 objects (n = 12) were assigned an active task, rendering them "task-relevant". The remaining 12 objects were merely passively present within the virtual apartment, and because no actions could be performed on them, they were defined as "task-irrelevant." Which specific objects were task-relevant versus task-irrelevant was pseudorandomized across participants. This ensured a balanced distribution of task-relevant and task-irrelevant items across both congruency conditions (congruent vs. incongruent) and across all three rooms. Additionally, the task order was pseudorandomized to avoid memory clustering, ensuring no more than two interactions per room or room category occurred after another. To complete each task, participants identified and approached the relevant object (e.g. the toaster with untoasted bread) and pressed the execution button for five seconds. This action would lift the object (e.g. slices of bread) with a loading bar indicating progress. After finishing (e.g. the toaster now appeared with toasted bread), participants could access new instructions by pressing the instruction button (see Fig. 1). The updates to the EVE task compared to previous versions involved drawing task-relevant items from the pool of all 24 objects (instead of using 12 predetermined objects) and adding a loading bar during interactions to enhance the sense of self-performed action.

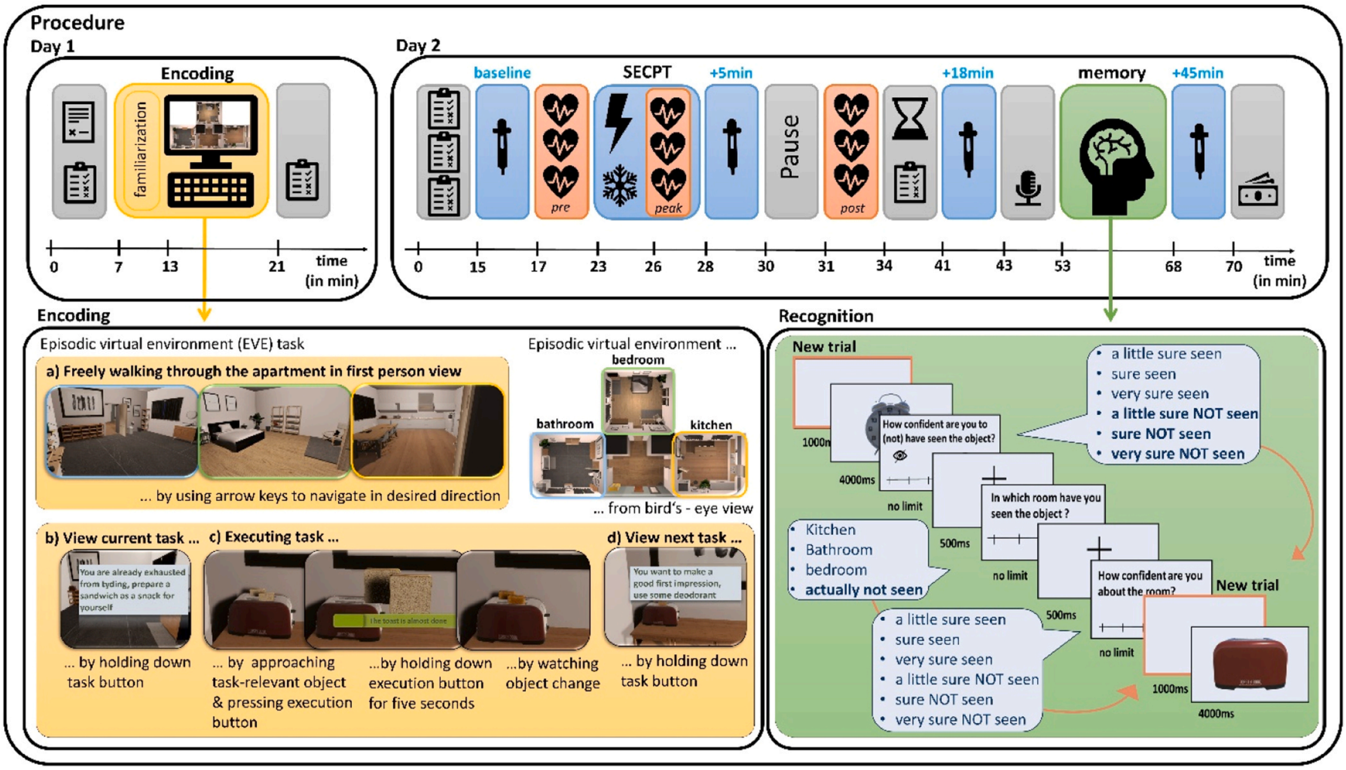

Fig. 1. Experimental timeline and task details. The figure illustrates the two-day experimental design, highlighting the sequence and duration of tasks. Icons with writing represent different questionnaires, the information sheet, and informed consent. On Day 1, participants underwent a familiarization phase followed by an encoding phase in a virtual environment using a desktop setup. On Day 2 (approximately 24 h after the encoding phase), after habituation, stress was induced using the Socially Evaluated Cold-Pressor Test (SECPT) or a control condition. Heart symbols indicate blood pressure and heart rate measurements (pre, peak, post stress). Salivary samples (blue) were collected at four time points (baseline, +5 min, +18 min, +45 min post-stress onset). Before debriefing, participants completed a free recall task (relevant to another project) and an old/new recognition task. Below the timeline, the section on the left shows the details of the encoding phase. The top right corner displays a bird's-eye view of the virtual apartment for illustration. Yellow background images represent participants' first-person perspective in the Episodic Virtual Environment (EVE) task, consisting of 12 tasks. Panels (a–d) illustrate an example task sequence. The right section below the timeline presents the details of the recognition task. Participants viewed 48 objects (24 targets, 24 lures) in randomized order and rated whether each object had been seen in the virtual environment (old/new decision) with a confidence scale (1–6). If identified as "old," participants selected the room (bedroom, bathroom, kitchen, or not seen) and rated their confidence in the room rating. New trials began immediately after "not seen" ratings or after corrections to "not seen at all" during the room rating phase.

2.3. Retrieval

We employed a classical old/new item recognition task, extended by an additional room rating choice and corresponding confidence rating (see Fig. 1, Zöllner et al., 2023). Participants viewed 48 household objects (24 targets, 24 lures). Each trial began with a 100 ms blank screen followed by a colored object on a grey background for 4000 ms. Recognition was indicated on a 6-point Likert scale (1 = very certainly not seen to 6 = very certainly seen, presented in German). For "new" responses (1−3), the next trial started immediately; for "old" responses (4−6), participants selected one of three rooms they had encountered during encoding (kitchen, bathroom, bedroom), with the object having appeared in one of them. Erroneous "old" ratings could be corrected before proceeding. Afterwards, participants rated their confidence in the room choice on another 6-point Likert scale (1 = very uncertain to 6 = very certain). A fixation cross (500 ms) was shown between rating screens. The task took ~15 min without response time limits.

To assess scenario construction during memory retrieval, we focused on incongruent objects only. For congruent objects, it is impossible to disentangle episodic from semantic memory in room recall, as both would point to the same response. In contrast, room responses to incongruent objects allow us to differentiate memory types more clearly. These responses were classified as either correct episodic retrieval (e.g., toaster in the bathroom) or errors: (a) semantic errors (toaster in the kitchen), or (b) unrelated errors (toaster in the bedroom) being neither consistent with the correct episode nor with the object's semantic category.

2.4. Stress induction, saliva sampling and analysis

In a randomized design, 88 participants (44 in the stress group: 22 women; 44 in the control group: 22 women) underwent the Socially Evaluated Cold-Pressor Test (SECPT; Schwabe et al., 2008) to induce psychological and physiological stress. Participants submerged their non-dominant hand in ice-cold water (0–2°C) for three minutes while being videotaped by an experimenter of the opposite sex. They were told that these video recordings would be later evaluated for facial expression. Throughout the task, they should look directly into the camera and keep their hand in the ice-cold water for as long as possible. Control participants used warm water (35–37°C) for the same duration without video-recorded observation. Post-exposure, participants rated the difficulty, stressfulness, painfulness, and unpleasantness of their experience on a scale from 0 (not at all) to 100 (very much) adopted from Schwabe et al. (2008). We recorded systolic (BPsys) and diastolic blood pressure (BPdia) and heart rate (HR) as a marker of SNS activity three times in a row, two minutes before (baseline), during (peak), and eight minutes after (post) stress onset, using an Omron M700 Intelli IT device (Omron Healthcare Co., Kyoto, Japan).

Saliva samples were collected at four different testing time points (baseline, +5 min, +18 min, +45 min; + indicating reference to onset of the stress/control condition) using Salivettes® (Sarstedt, Nümbrecht, Germany) to measure alpha-amylase (as an additional index of SNS activity) and free cortisol concentrations (as a measure of HPA axis activity), along with affect ratings (using the German version of the Positive Affect Negative Affect Schedule, PANAS (Breyer & Bluemke, 2016)). The saliva samples were stored at −20 °C until analyzed with a commercial enzyme-linked immunosorbent assay (IBL International, Hamburg, Germany) in our in-house biochemical laboratory, with inter- and intra-assay coefficients of variation below 10%. A colorimetric test was applied, using 2-chloro−4-nitrophenyl-α-maltrotriosoide (CNP-G3) as a substrate reagent to measure salivary alpha-amylase concentrations (Lorentz et al., 1999). Intra- and inter-assay coefficients of variability were below 6.74%.

2.5. Experimental design and general procedure

We used a 2 × 2 design with the between-subjects factors stress (stress vs. control) and sex (women vs. men). Participants were randomly assigned to stress ($n = 44$) or control ($n = 44$) groups (22 women and 22 men each). The study was conducted at Ruhr University Bochum between 9 a.m. and 6 p.m., following evidence of stress-induced retrieval deficits throughout the day (Smeets, 2011). Participants were assigned to a testing time window either in the morning (9 a.m. - 1 p.m.) or in the afternoon (1 p.m. - 6 p.m.) balanced by group and sex.

On the first day, participants engaged in an encoding task in a computer-based virtual environment via Unity® software, followed by the Igroup Presence Questionnaire (IPQ) to evaluate their sense of presence (Schubert et al., 2001). On the second day (approximately 24 h later), we assessed a demographic questionnaire, including video game experience and mental imagery scores from the vividness of visual imagery Questionnaire (VVIQ; Marks, 1973) and the Brief Symptom Inventory (BSI; Franke, 2000) as filler tasks during the habituation phase. We induced acute stress 30 min before a free recall task [1] (relevant for another project) and before a modified item recognition task took place to assess memory for individual objects and the spatial contexts (i.e., rooms) in which they had been encoded. All tasks occurred in the same chamber on a computer, while stress induction was realized in a separate chamber. Participants received 25€ for reimbursement and were debriefed.

2.6. Statistics and data preparation

Recognition, PANAS and IPQ data were extracted from computer-generated log-files using MATLAB (The MathWorks Inc, 2021) including the Open Toolbox for Behavioral Research (Otto & Rose, 2023); other questionnaire data were PDF based. We prepared the behavioral data for further analysis using Python 3.8 in the Spyder environment (Raybaut, 2009; Van Rossum & Drake Jr., 2009). All statistical analyses were conducted in R Studio (RStudio Team, 2019) with R (R Core Team., 2020) version 4.3.1. A significance level of $\alpha = .05$ was set and the data were checked for normality using Shapiro-Wilk tests and log-transformed if necessary. Additionally, we examined all dependent variables for homogeneity of variance with Levene's test and reported Greenhouse-Geisser corrected p-values and degrees of freedom in cases where sphericity was violated. Significant interactions were solved using appropriate (Tukey-corrected) post-hoc tests. Partial eta squared (η_p^2) values were provided as estimates of effect sizes. All analyses of variance (ANOVA) included the between-subject factors stress (stress vs. control) and sex (men vs. women). For group comparisons of age and body mass index (BMI), two separate 2 × 2 ANOVA were used.

To verify successful stress induction, we conducted separate mixed-design ANOVA for various measurements. For salivary cortisol and alpha-amylase, we applied the repeated measures factor time with the factor levels baseline vs. + 5 min vs. + 18 min vs. + 45 min, whereas for BPsys, BPdia and HR, the repeated measures factor time comprised the factor levels baseline vs. peak vs. post-stress. For the difficulty, stressfulness, painfulness, and unpleasantness ratings after the stress/control condition, we conducted separate 2 × 2 ANOVA.

To ensure that participants executed the recognition task correctly (and did not randomly press the same button to speed up the task), we calculated d', namely the standardized difference between hit rate and false alarm rate, in line with the method outlined by Macmillan and Creelman (1996), as a measure of correct discrimination between targets and lures. To investigate the discriminatory ability d' between groups, we conducted a t-test. To check the effect of stress and sex on memory performance, we conducted two separate ANOVA, with the dependent variables 'number of congruent target objects' and 'number of incongruent target objects'.

To ensure comparability with a previous study (Zöllner et al., 2023) and to investigate the overarching factors influencing memory accuracy, we explored how the objects' characteristics, such as congruency and task-relevance as well as stress and sex, predicted correct room recall as a trial-level binary variable using a logistic linear mixed model (LMM) analysis. This analysis was utilized for this step because it is specifically designed to handle binary dependent variables at the individual trial level without requiring data aggregation, while simultaneously accounting for the nested, repeated-measures structure of the data by treating subjects as random factors. 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. This approach allowed us to account for both individual subject effects and object effects by including them in our model as random factors.

To analyze memory retrieval data with regard to semantic construction, we calculated the proportion of incongruent objects classified into each response category based on the total number of recognized incongruent objects. Responses regarding incongruent objects were classified as either correct episodic (e.g., toaster in the bathroom), semantic errors (toaster in the kitchen) or unrelated errors (toaster in the bedroom). We calculated the proportion of correct episodic responses, semantic errors, and unrelated errors for each participant by dividing the number of objects recalled in each category by the total number of incongruent objects recognized as 'old'. Because the standard LMM architecture is restricted to binary or continuous outcomes, analyses investigating the distribution of the three distinct, nominal response categories for incongruent objects (correct episodic responses, semantic errors, and unrelated errors) and effects of stress and sex were evaluated using mixed-design ANOVA on aggregated subject proportions.

Exploratorily, we investigated if the confidence levels for correct episodic responses, semantic errors and unrelated errors varied based on stress and sex by using separate ANOVA. Besides, we used correlation analyses between the mean values of the sense of presence experienced in the virtual environment (IPQ), experience with video games (demographic questionnaire), mental imagery scores (VVIQ) and the semantic construction score. The semantic construction score was calculated for each participant as the total number of recognized incongruent objects recalled in the semantically related room minus the total number of recognized incongruent objects recalled in the unrelated room.

3. Results

3.1. Sample description

We tested 88 participants in total aged between 18 and 34 years. Due

[1] During free recall, participants verbally recounted their memories of the virtual environment episode while being audio-recorded. To guide their focus, we first presented an example recording emphasizing objects and their locations, ensuring participants provided the detailed information needed to examine scenario construction.

to noncompliance with our inclusion criteria (antibiotic intake, intervening with circulating cortisol levels), we had to exclude one woman from the control group from all further analyses, resulting in 87 participants (43 women). The two separate ANOVA revealed a significant main effect of sex on age ($F_{(1,83)}$= 4.48, p = .037, η_p^2= .04) and BMI ($F_{(1,83)}$= 21.81, p < .001, η_p^2= .13). Men were slightly older and had a slightly higher BMI (mean age±SD: 24.30 ± 3.92 years; BMI: 24.57 ± 2.47 kg/m^2) than women (mean age±SD: 22.93 ± 4.19 years; BMI: 22.20 ± 2.24 kg/m^2). No significant effects of stress were found (both *ps* > .05), nor was there an interaction effect between sex and stress groups for age and BMI (both *ps* > .05), indicating that stress and control group participants did not differ in these measures.

3.2. Stress induction

3.2.1. Cortisol concentrations and alpha-amylase

For analyses of salivary samples, we additionally excluded three participants missing either two salivary measurements (n = 1) or the last salivary measurement (n = 2), resulting in 42 participants in the stress and 42 participants in the control group (see Table 1).

Over time, stress led to an increase in salivary cortisol (main effect of time: $F_{(1117)}$= 16.09, p < .001, η_p^2= .0002; main effect of stress: $F_{(1,80)}$= 10.27, p = .002, η_p^2= .11; Stress x Time interaction: $F_{(1117)}$= 43.33, p < .001, η_p^2= .10), compared to the control manipulation, indicating successful stress induction. Post-hoc *t*-tests confirmed that the groups neither differed in cortisol at baseline, nor directly after the SECPT (both *ps* > .05). However, 18 min after the stress onset and directly before the recognition task, stressed participants exhibited higher cortisol concentrations ($t_{(119)}$= 5.88, p < .001) than controls, which persisted still 45 min after stress onset ($t_{(119)}$= 5.51, p < .001). No further main or interaction effects occurred (all *ps* > .05).

The repeated measure ANOVA for alpha-amylase revealed a main effect of time only ($F_{(3217)}$= 4.98, p = .003, η_p^2= .01). Follow up *t*-tests revealed that alpha-amylase decreased from 5 min to 18 min after stress onset ($t_{(83)}$= 3.88, p < .001), while it increased from 18 min to the end of the experiment ($t_{(83)}$= 3.33, p = .001). We found no other significant differences between timepoints (all other *ps* > .05). No further main or interaction effects emerged (all *ps* > .05).

Table 1
Mean (±SEM) salivary cortisol concentrations (A), salivary alpha-amylase levels (B), negative affect ratings (C) and positive affect ratings (D) at baseline, after 5 min, 18 min, and 45 min after onset of the stress/control intervention. Data is shown separately for women and men in the stress and control group, respectively. The statistics are described in detail in the text.

	stress women (n = 22)	men (n = 20)	control women (n = 21)	men (n = 21)
(A) salivary cortisol (nmol/l)				
baseline	5.68 ± 5.33	5.33 ± 3.72	4.96 ± 3.52	5.11 ± 2.69
5 min after stress onset	5.06 ± 4.21	4.96 ± 3.38	4.28 ± 2.77	4.78 ± 2.69
18 min after stress onset***	9.04 ± 7.36	10.20 ± 5.41	3.96 ± 2.31	4.25 ± 2.42
45 min after stress onset***	7.89 ± 5.71	6.82 ± 3.98	3.16 ± 1.65	3.32 ± 2.39
(B) alpha-amylase (U/ml)				
baseline	169.09 ± 93.54	148.80 ± 106.57	156.70 ± 109.37	165.06 ± 126.84
5 min after stress onset	188.66 ± 103.90	184.94 ± 153.81	160.92 ± 124.16	151.80 ± 96.72
18 min after stress onset	154.98 ± 88.52	179.63 ± 175.02	141.44 ± 90.85	143.12 ± 92.69
45 min after stress onset	175.04 ± 104.76	192.22 ± 195.42	159.34 ± 92.57	158.13 ± 94.76
(C) negative affect ratings				
baseline	3.59 ± 2.34	4.35 ± 2.41	3.86 ± 2.41	2.67 ± 1.65
5 min after stress onset	4.68 ± 3.47	5.25 ± 3.14	2.43 ± 1.86	2.57 ± 1.80
18 min after stress onset	3.23 ± 2.69	3.00 ± 2.13	2.05 ± 1.86	2.57 ± 2.06
45 min after stress onset	1.91 ± 1.44	2.60 ± 1.98	1.38 ± 0.80	2.19 ± 1.57
(D) positive affect ratings				
baseline	6.45 ± 4.00	6.30 ± 3.25	7.81 ± 4.17	5.90 ± 3.27
5 min after stress onset	7.77 ± 3.37	8.50 ± 4.07	6.81 ± 3.34	7.67 ± 3.86
18 min after stress onset	6.95 ± 3.47	6.80 ± 3.76	7.48 ± 3.46	6.90 ± 3.83
45 min after stress onset	7.59 ± 4.10	7.00 ± 3.09	7.43 ± 3.26	7.76 ± 4.59

*** p < .001 (*t*-tests), significant difference between the stress and control group.

3.2.2. Cardiovascular measurements

3.2.2.1. Blood pressure. For systolic and diastolic blood pressure, the two separate repeated measure ANOVA revealed significant main effects of time [BPsys: $F_{(2153)}$= 30.57, p < .001, η_p^2= .001 and BPdia: $F_{(2141)}$= 32.07, p < .001, η_p^2= .01] and stress [BPsys: $F_{(1,83)}$= 9.91, p = .002, η_p^2= .11 and BPdia: $F_{(1,83)}$= 10.31, p = .002, η_p^2= .11] as well as a significant Time × Stress interaction [BPsys: $F_{(2153)}$= 89.14, p < .001, η_p^2= .15 and BPdia: $F_{(2141)}$= 66.73, p < .001, η_p^2= .23]. Post hoc tests revealed that blood pressure between groups only differed during peak measurement, during which the stress group showed higher systolic and diastolic blood pressure compared to the control group [$t_{(103)}$= 6.29, p < .001 and $t_{(149)}$= 8.65, p < .001 respectively; see Table S1]. No differences occurred at baseline or post-stress (*ps* > .05).

Additionally, a significant main effect of sex ($F_{(1,8)}$= 53.46, p < .001, η_p^2= .39) and a significant Time × Stress × Sex interaction ($F_{(2,13)}$= 6.10, p = .003, η_p^2= .01) was found for systolic blood pressure only (for diastolic blood pressure all *ps* > .05). Men in the stress group consistently exhibited higher BPsys than women in both the stress and control groups at baseline [(stress: ($t_{(110)}$= 5.81, p < .001), control: ($t_{(110)}$= 5.52, p < .001)], at peak [(stress: ($t_{(110)}$= 3.87, p = .010), control: ($t_{(110)}$= 10.12, p < .001)] and at the post measurement time point [stress: ($t_{(110)}$= 5.25, p < .001), control: ($t_{(110)}$= 4.84, p < .001)]. Similarly, men in the control group exhibited significantly higher BPsys than women in the control group at baseline ($t_{(110)}$= 4.71, p < .001), peak ($t_{(110)}$= 5.35, p < .001) and post measurement ($t_{(110)}$= 3.86, p = .010). When comparing men in the control group to women in the stress group, this difference was significant only at baseline ($t_{(110)}$= 4.97, p < .001) and post measurement ($t_{(110)}$= 4.25, p < .003). However, women in the stress group exhibited a pronounced BPsys increase during the SECPT, resulting in a temporary absence of sex differences in BPsys at the peak measurement (p > .05).

In summary, the induction of stress successfully increased both systolic and diastolic blood pressure. Diastolic blood pressure was unaffected by sex. Systolic blood pressure was affected by stress and sex. Men exhibited higher systolic blood pressure than women, except during the SECPT, where women in the stress group and men in the control group displayed similar levels.

3.2.2.2. Heart rate. The repeated measure ANOVA revealed a significant main effect of time ($F_{(2,13)}$= 43.10, p < .001, η_p^2= .01) and a Time × Stress interaction ($F_{(2,13)}$=24.11, p < .001, η_p^2=.05). Post hoc *t*-tests indicated a significantly higher heart rate in the stress compared to the control group during the SECPT only ($t_{(113)}$= 3.29, p = .001, baseline and post-stress: *ps* > .05; see Table 2). No further significant main or interaction effects emerged. In summary, stress induction successfully

Table 2
Mean (±SEM) heart rate (A), diastolic blood pressure (B), systolic blood pressure (C) at baseline, peak and post after onset of the stress/control intervention. Post SECPT ratings (D) were obtained directly after the stress/control intervention. Data is shown separately for women and men in the stress and control group, respectively. The statistics are described in detail in the text.

	stress		control	
	women (n = 22)	men (n = 22)	women (n = 21)	men (n = 22)
(A) heart rate (bpm)				
baseline	71.91 ± 11.59	64.73 ± 11.77	67.02 ± 8.07	65.71 ± 9.91
peak*	79.23 ± 14.58	75.65 ± 10.81	69.55 ± 7.34	68.86 ± 10.45
post	69.61 ± 10.64	63.07 ± 10.84	69.06 ± 6.91	67.67 ± 8.97
(B) diastolic blood pressure (mmHg)				
baseline	63.23 ± 5.08	66.03 ± 9.45	65.41 ± 6.24	65.39 ± 6.86
peak***	78.88 ± 6.71	85.33 ± 9.89	64.06 ± 6.63	63.32 ± 7.40
post	68.80 ± 4.74	69.83 ± 7.64	69.29 ± 8.61	67.81 ± 7.88
(C) systolic blood pressure (mmHg)				
baseline	99.65 ± 9.02	118.09 ± 11.02	99.38 ± 10.49	113.23 ± 6.88
peak***	116.92 ± 12.31	139.65 ± 12.02	97.52 ± 10.63	110.39 ± 5.24
post	100.85 ± 8.03	119.25 ± 8.54	101.68 ± 10.18	113.92 ± 5.83
(D) post SECPT ratings				
difficult***	79.09 ± 2.71	62.27 ± 6.48	2.86 ± 1.71	6.82 ± 2.97
unpleasant***	85.00 ± 3.00	63.18 ± 6.36	7.62 ± 4.57	7.73 ± 3.72
stressful***	72.27 ± 4.21	55.46 ± 5.49	5.71 ± 2.72	8.64 ± 3.91
painful***	87.27 ± 2.39	66.82 ± 5.24	0.95 ± 0.95	1.36 ± 0.75

* $p < .05$, *** $p < .001$ (t-tests), significant difference between the stress and control group.

increased heart rate.

3.2.3. Ratings

3.2.3.1. Affect. We acquired negative affect ratings parallel to cortisol measurements using the PANAS. The repeated measures ANOVA revealed a significant main effect of time ($F_{(3215)}$= 17.93, $p < .001$, η_p^2= .0003) and a significant interaction effect of Stress x Time ($F_{(3215)}$= 5.69, $p < .001$, η_p^2= .03). A post hoc Welch test showed that the stress group indicated a decrease in negative affect over time, except for immediately after the intervention compared to baseline ($p < .05$). At this point, their negative affect was higher than indicated by the control group ($t(150) = 3.78$, $p = .006$). No differences in negative affect between the groups occurred at any other time (all other $ps > .05$). Likewise, self-reported stress responses did not significantly differ between women and men ($p > .05$). No further main or interaction effects emerged (all $ps > .05$).

For reasons of completion, we exploratorily analyzed the positive affect in the same manner. The repeated measures ANOVA revealed a main effect of time ($F_{(3207)}$= 4.28, $p = .006$, η_p^2= .0005). Post hoc pairwise comparisons revealed a significant decrease in positive affect from directly after the intervention to 18 min. after its onset ($t_{(83)}$= 3.50, $p < .001$). Besides, we found a main effect of stress ($F_{(1,80)}$= 4.46, $p = .038$, η_p^2= .05). The positive affect in the stress group was generally higher than in the control group. However, we found neither a Stress x Time interaction ($p > .05$) nor other main or interaction effects (all $ps > .05$). In summary, stress did not affect positive affect over time, but effectively increased negative affect in the stress group only.

3.2.3.2. SECPT ratings. The separate ANOVA revealed significant main effects of stress [difficult: $F_{(1,83)}$= 211.26, $p < .001$, η_p^2= .49; unpleasant: $F_{(1,83)}$= 213.53, $p < .001$, η_p^2= .50, stressful: $F_{(1,83)}$= 193.15 $p < .001$, η_p^2= .50; and painful: $F_{(1,83)}$= 987.94, $p < .001$, η_p^2= .84]. The stress group rated the SECPT as more difficult, stressful, painful, and unpleasant compared to the control group. Additionally, a significant Sex × Stress interaction effect occurred for the rating of difficulty ($F_{(1,83)}$= 4.47, $p = .037$, η_p^2= .05) and painfulness ($F_{(1,83)}$= 4.47, $p = .037$, η_p^2= .05). In the control group, there was no difference between men and women (both $ps > .05$), while in the stress group women rated the situation as harder ($t_{(82)}$= 9.11, $p < .001$) and more painful ($t_{(82)}$= 20.13, $p < .001$) than men. No other main effect was significant (all $ps > .05$). In summary, the intervention led to increased overall stress ratings in the stress compared to the control group.

Taken together, the stress induction was successful.

3.3. Recognition task

3.3.1. Memory performance for targets (congruent and incongruent) under stress

In the initial phase of our analysis, we assessed whether recognition memory sensitivity differed significantly between groups. The t-test comparing the sensitivity measure (d') between the control (1.92 ± 0.76; M±SD) and the stress group (2.00 ± 0.79; M±SD) showed no significant difference ($t_{(85)}$= 0.50, $p = .621$). In addition to our preregistration, an ANOVA revealed no significant differences based on stress or any interaction with sex (all $ps > .05$).

Then, we assessed memory performance during the recognition task. Participants recognized an average of 18 out of 24 objects (17.8 ± 2.43; M±SD) with about nine congruent (8.83 ± 1.48; M±SD) and nine incongruent objects (8.98 ± 1.52; M±SD) out of 12 respectively. Two separate ANOVA (for congruent and incongruent objects) indicated no differences in recognized objects based on stress or sex (all $ps > .05$). In summary, the recognition task was conducted successfully, as revealed by positive d' values, but we found no differences in recognition memory performance concerning stress and sex.

3.3.2. Prediction of correct room recall by object characteristics and stress

We assessed the influence of the task object characteristics (congruency, task-relevance), stress and sex on memory performance by fitting a logistic linear mixed model to the trial-level recognition task data. To account for within-subject correlation in the data, 'subject' was included as a random effect, allowing for variability in intercepts. While we originally planned to include 'object' as an additional random factor to model object-specific variance, we ultimately chose to omit this factor due to the model failing to converge when both random factors were included. This analysis revealed a significant effect of the predictor congruency (OR=7.42, 95% CI (OR) [3.44; 16.02], $p < .001$), with correct room recall being significantly more likely for congruent as opposed to incongruent objects. Moreover, there was a significant effect of the predictor task-relevance (OR=6.48, 95% CI (OR) [3.38; 12.41], $p < .001$), indicating that task-relevant objects were significantly more likely to be recalled in the correct room. In addition to these main effects, the interaction between the predictors congruency and task-relevance was significant (OR=0.27, 95% CI (OR) [0.10; 0.73], $p = .010$, Fig. 2). Post-hoc comparisons of estimated marginal means showed that there was a higher difference of the predicted probabilities of correct room recall between congruent and incongruent *task-irrelevant* objects (OR=7.09, 95% CI_{Tukey} (OR) [4.78; 10.52], p_{Tukey}< .001) than between congruent and incongruent *task-relevant* objects (OR=3.44, 95% CI_{Tukey} (OR) [2.35; 5.03], p_{Tukey}< .001). Furthermore, there was a higher difference in the predicted probabilities of correct room recall

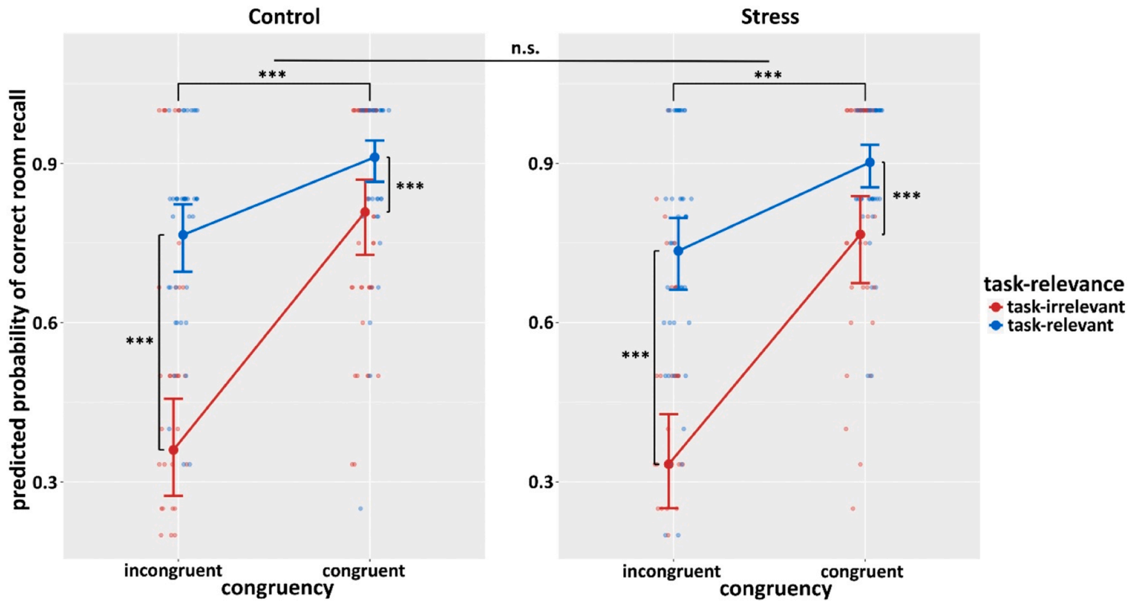

Fig. 2. Predicted probabilities (marginal effects) of correct room recall in the recognition memory task per levels of the predictors congruency and task-relevance; separately for the control group (left panel) and the stress group (right panel). Results indicated that correct room recall was significantly more likely for congruent as opposed to incongruent objects. In addition, task-relevant objects were significantly more likely to be recalled in the correct room. The interaction between the predictors congruency and task-relevance revealed a higher difference of the predicted probabilities of correct room recall between congruent and incongruent *task-irrelevant* objects than between congruent and incongruent *task-relevant* objects. Moreover, a higher difference occurred for the predicted probabilities of correct room recall between task-relevant and task-irrelevant *incongruent* objects than between task-relevant and task-irrelevant *congruent* objects. Error bars represent 95% confidence intervals of the predicted probabilities. Single data points represent individual predicted probabilities for each participant and each combination of predictor levels. *** $p < .001$.

between task-relevant and task-irrelevant *incongruent* objects (OR=5.76, 95% CI_{Tukey} (OR) [4.11; 8.05], $p_{Tukey} < .001$) than between task-relevant and task-irrelevant *congruent* objects (OR=2.80, 95% CI_{Tukey} (OR) [1.82; 4.30], $p_{Tukey} < .001$). The model did not indicate significant main effects of stress or sex or any interactions involving these predictors on the probability of correct room recall (Fig. 3).

Overall, these results suggest that recall of the room an object was encountered in was superior for objects that the participant interacted with during encoding and for congruently placed objects. Memory for the original room was especially error-prone for task-irrelevant incongruently placed objects compared to task-irrelevant congruently placed objects, suggesting that episodic memory for task-irrelevant objects and their spatial contexts was weak and that recall of objects' associated rooms may have relied primarily on semantic construction.

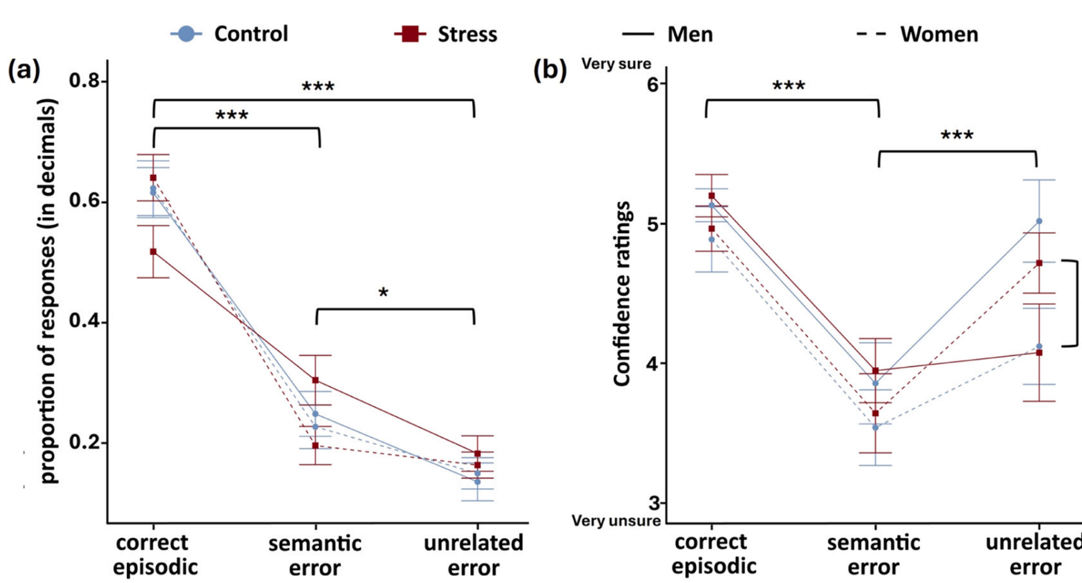

Fig. 3. (a) Mean (±SEM) proportion of correct episodic responses, semantic errors, and unrelated errors, separately displayed as a function of stress and sex. Results revealed a higher proportion of correct episodic responses compared to semantic errors and unrelated errors as well as a higher proportion of semantic errors compared to unrelated errors. (b) Mean (±SEM) confidence ratings for correct episodic responses, semantic errors, and unrelated errors, separately displayed as a function of stress and sex. Results indicated that the confidence for semantic errors was significantly lower compared to correct episodic responses and unrelated errors. Additionally, women had higher confidence ratings for unrelated errors compared to men in the stress, but not in the control group. * $p < .05$, *** $p < .001$.

3.3.3. Semantic construction and stress

To investigate the memory systems involved, we focused on the memory for incongruent objects. The mixed ANOVA revealed a significant main effect of response type ($F_{(2149)}$= 109.60, $p < .001$, η_p^2= .42). Post-hoc pairwise comparisons indicated a significantly higher proportion of correct episodic responses (0.60±0.20; $M \pm SD$) compared to semantic errors (0.24 ± 0.17; $M \pm SD$; $t_{(255)}$= 13.35, p_{Tukey}< .001) and unrelated errors (0.16 ± 0.13; $M \pm SD$; $t_{(255)}$= 16.96, p_{Tukey}< .001). Additionally, there was a higher proportion of semantic errors compared to unrelated errors ($t_{(255)}$= 3.61, p_{Tukey}= .020). No other main or interaction effects were significant (all $ps > .05$).

In summary, participants demonstrated intact episodic retrieval by accurately identifying the original room for a majority of the items. Furthermore, when episodic memory failed, incorrect responses were not distributed randomly but were significantly more likely to be substituted with the semantically congruent room than with an unrelated room, confirming a systematic pattern of semantic construction. However, neither episodic memory nor semantic construction was significantly affected by acute stress or sex.

3.3.4. Confidence ratings

To investigate the differences in confidence for the distinct response types (i.e., correct episodic responses, semantic errors, unrelated errors), we excluded all data of participants with zero responses in one of these bins, resulting in a sample size of $n = 53$ (stress: $n = 29$, control: $n = 24$). The mixed ANOVA revealed a significant main effect of response type ($F_{(2,98)}$= 35,39 $p < .001$, η_p^2= .09). Post-hoc pairwise comparisons indicated that the confidence for correct episodic responses (4.96 ± 0.87; $M \pm SD$) was significantly higher than for semantic errors (3.59 ± 1.07; $M \pm SD$; $t_{(153)}$= 6.61, p_{Tukey}< .001), but did not differ from the confidence for unrelated errors (4.54 ± 1.22; $M \pm SD$; $p > .05$). Thus, the confidence for unrelated errors was also significantly higher than confidence for semantic errors ($t_{(153)}$= 4.59, p_{Tukey}< .001).

Additionally, we found a significant Group x Sex x Response Type interaction ($F_{(2,98)}$= 4.94, $p = .009$, η_p^2= .04). Separate post hoc ANOVA showed no significant effects of acute stress and sex on confidence levels of correct episodic responses or semantic errors (all $ps > .05$). However, there was a significant interaction between stress and sex concerning the confidence levels of unrelated errors ($F_{(1,49)=}$7.28, $p = .010$, η_p^2 = .13). Post hoc tests revealed that only in the stress group confidence differed according to sex ($t_{(19.35)}$= 2.31, $p = .032$), with women (5.07 ± 0.63; $M \pm SD$) having higher confidence ratings than men (4.12 ± 1.45; $M \pm SD$) for unrelated errors, while no differences occurred in the control group. Furthermore, women in the stress group displayed significantly greater confidence in their unrelated errors compared to women in the control group (4.15 ± 1.24; $M \pm SD$), as confirmed by an additional t-test ($t_{(19.25)}$= 2.47, $p = .023$). No other main or interaction effects were significant (all $ps > .05$). In summary, the overall three-way interaction seems to be driven by differences in higher confidence for unrelated errors, specifically among women under stress.

3.4. Correlation of individual propositions on semantic construction

We applied Pearson correlation analyses to investigate the link between the subject-specific degree of semantic construction in the recognition task and the self-reported sense of presence in the virtual environment (assessed via the IPQ questionnaire) and the self-reported mental imagery score (assessed via the VVIQ questionnaire). The correlation between the semantic construction scores and the IPQ-derived sense of presence scores was not statistically significant ($r(83) = 0.08$, 95% CI [−0.14; 0.29], $p = .469$). The same was true for the correlation between the semantic construction scores and the VVIQ-derived mental imagery scores ($r(86) = -0.06$, 95% CI [−0.27; 0.15], $p = .581$). In contrast to the preregistration, we could not perform a correlation analysis for participants' experience with video games. This was due to a large proportion of our sample not reporting any experience with video gaming and the remaining responses to our questionnaire being very vague and irregular, rendering objective quantification of video gaming experience impossible.

4. Discussion

The current study investigated how acute stress impacts semantic construction when remembering past events, exploratorily examining sex-related differences and confidence ratings. We aimed to enhance the understanding of modifying factors such as congruency, task-relevance, and stress in predicting correct episodic memories. Participants were more likely to commit semantic errors than unrelated errors by systematically choosing the semantically associated room over an unrelated room when episodic memory failed, though acute stress did not change this effect. Confidence levels were highest for correct episodic responses, followed by unrelated and semantic errors. Under stress, women showed greater confidence in unrelated errors, while their semantic knowledge use remained stable. Generally, congruent objects that were actively engaged with were recalled with the highest accuracy. Altogether, these findings support a constructive account of episodic memory retrieval which involves completion by semantic information. This research builds on previous studies by incorporating confidence metrics and emphasizing encoding factors that may affect memory, including object congruency and task-relevance.

4.1. Semantic construction and stress effects

Our findings suggest that when participants made an error regarding the spatial context (i.e., the room) in which an incongruent object was encoded, their incorrect choices often relied on knowledge about the object and its semantically related room category, rather than on random, unrelated information. Importantly, regarding committed semantic errors (e.g., remembering a toaster in the kitchen, although it was originally placed in the bathroom), it remains unclear whether episodic details were entirely absent and thus replaced by semantic memory, or whether semantic information was simply more easily accessible compared to weak episodic memories. Therefore, we interpret these errors not as a matter of either episodic or semantic memory dominating retrieval entirely, but rather as evidence for their interaction. We assume that episodic memory processes are preserved to some extent - for instance, remembering the object itself and that it was encountered in the virtual environment (the gist of the scenario) - while missing details are substituted with semantic memory, leading to a systematic bias toward the semantically congruent room selection over an unrelated one. Regardless of whether this construction reflects the absence of episodic detail or its suppression by semantic memory, our results highlight the interplay of episodic and semantic processes during scenario construction in a realistic setting and after a long retrieval interval. This interpretation aligns with prior work on semantic construction (Klein et al., 2025; Zöllner et al., 2023) and the Scenario Construction model (Cheng et al., 2016; Cheng, 2024).

While we interpret our findings through the lens of the Scenario Construction Framework (Cheng et al., 2016; Cheng, 2024), we acknowledge alternative theoretical accounts. During encoding, the EVE task presented participants with a complex, naturalistic episode involving spatial navigation, temporal action sequences, and object manipulation. According to the Scenario Construction Framework, such an experience results in the storage of sparse episodic traces that capture the gist of the event. During retrieval, these traces interact with semantic information to reconstruct a coherent scenario (Cheng et al., 2016; Cheng, 2024). Because our encoding task was highly generative and spatially extended, we argue that retrieving the spatial context of an object likely involves reconstructing the broader scenario rather than simply accessing a singular associative tag. However, because our retrieval phase utilized a multiple-choice recognition and room-recall format, the observed response patterns – correct retrieval, semantic errors, and unrelated errors – could also be interpreted through mechanisms of item-context associative memory (Staresina et al., 2009; van Kesteren et al., 2012, Yonelinas et al., 2019). In such models, the semantic congruence between an item and its context facilitates stronger associative binding at encoding, and errors during retrieval are viewed as semantic interference or misattributions rather than constructive scene generation. Future studies utilizing generative recall tasks, such as asking participants to freely reconstruct the temporal sequence of actions and their precise spatial locations as well as broader spatiotemporal context, are needed to further disentangle scenario construction from item-context associative memory.

Although we induced stress successfully, there was no decline in episodic memory performance across groups, despite prior results (Buchanan et al., 2006; Shields et al., 2017; however, stress was also shown to increase memory recognition of faces, see e.g. Pötzl et al., 2023). Therefore, we cannot support the hypothesis that acute stress weakens episodic memory, leading to increased semantic construction. No stress effect occurred for the proportion of correct episodic responses or differentiation between targets and lures (d') across groups. This null finding aligns with previous results, which indicated no effect of cortisol administration on recognition memory and correct room recall either (and consequently no influence on corresponding semantic construction), in the same task as used here (Klein et al., 2025). Furthermore, the lack of a stress effect on semantic construction may relate to the used recognition task, which seems to be less vulnerable to impairing stress effects (Gagnon & Wagner, 2016). Retrieval of neutral stimuli (as we used in the current study) also appears to be less affected by stress (Wolf, 2009). Additionally, the nature of the stressor itself might play a critical role. While the SECPT triggers a robust physiological response, it lacks the intense social-evaluative threat and unpredictability that characterize other paradigms like the Trier Social Stress Test (TSST; Kirschbaum et al., 1993; see Klier & Buratto, 2020). Previous research has demonstrated that the TSST can significantly impair both item and associative recognition memory (Guez et al., 2016). Thus, it is possible that a stronger, purely psychosocial stressor might be more effective in disrupting episodic retrieval and shifting the balance toward semantic construction in this paradigm. Finally, our recognition task took place 30 min after stress onset during the expected cortisol peak. Future research might aim to realize the recognition task earlier or even during acute stress to also include the potential contribution of noradrenergic effects on semantic construction.

4.2. Confidence ratings of semantic construction

Our finding of the highest confidence levels in correct episodic responses, compared to lower confidence in semantic errors, aligns with previous research indicating that recollection is typically linked to high confidence in recognition, while familiarity is associated with a more variable degree of confidence (Yonelinas, 2001). We acknowledge that we cannot claim with certainty whether the observed pattern of confidence for semantic errors compared to correct episodic responses and unrelated errors reflects familiarity, guessing, or a pattern typical of scenario construction. Nevertheless, a significantly higher proportion of semantic errors occurred instead of an equal distribution between semantic and unrelated room choices, which might be assumed if errors were distributed randomly, and which thus speaks against guessing at least. We eliminated the guessing aspect by allowing participants to revise their claims about remembering seeing the object by offering the option "actually not seen at all" during room ratings. Finally, we removed any time pressure during responses. Together, this approach speaks at least against guessing and, thus, rather for familiarity or a reconstructive pattern. Future research could explore whether room responses indicate familiarity, guessing, or incorrectly constructed memories based on prior knowledge. This could be achieved by incorporating phenomenological responses (remember, know, just familiar, and guess), similar to Coane et al. (2021) or Umanath and Coane (2020). Additionally, researchers could ask participants if they can vividly visualize the object in the chosen room and specify exactly where in the room they remember the object was located. This would help determine whether participants simply guessed or if incorrect details were actually integrated into their constructed scenarios.

Interestingly, the confidence in unrelated errors was significantly higher than confidence in semantic errors, even though both are technically memory errors for which confidence would be expected to be low. However, the increase in confidence regarding erroneous recollections corresponds with previous research documenting elevated confidence levels associated with source misattribution (Schacter, 1999; Zaragoza & Lane, 1994). Specifically, stress appears to increase confidence in unrelated errors in women. However, the underlying mechanism by which stress enhances confidence while not influencing behavioral indicators that favor semantic over unrelated errors remains unclear, particularly given that this effect was observed exclusively in women. It is noteworthy that in women experiencing acute stress, the confidence levels associated with erroneous memories rise to those of episodic memories. Numerous studies on false memories have indicated that participants exhibit greater confidence in genuine memories they have actually experienced than in those they were merely led to believe they had experienced (Loftus, 2005). Thus, one could interpret our finding that stressed women perceive a memory when missing details are substituted with random information as a genuine memory, suggesting a constructive nature of memories during scenario construction. Our findings suggest that stress prior to retrieval may exacerbate difficulties in distinguishing between accurate and different types of false memories based on levels of confidence in women.

4.3. Prediction of correct room recall by object characteristics and stress

We replicated the findings from previous studies that participants had a better ability to recall the room where an object was encountered if they had interacted with the object during the encoding process (Klein et al., 2025; Zöllner et al., 2023). Our results show that semantic congruency of objects during encoding promotes correct room recall, i.e., accurate episodic memory. Compatible with this observation are various studies on the semantic congruency effect that found enhanced subsequent long-term recognition memory of words after semantic congruency at encoding but also that information congruent with existing schemas is recognized and recollected more accurately and more quickly relative to incongruent information (Brod & Shing, 2019; Greve et al., 2007; van Kesteren et al., 2010). Memory for the original room was particularly susceptible to errors with incongruently compared to congruently placed non-interaction objects. Similarly, a multisensory VR study found that participants remembered significantly more objects with congruent odors than those with incongruent odors or no odors at all. Interaction type did not significantly affect recall, even though interaction with congruent objects was rated as more pleasant (Rantala et al., 2024). However, interacting with task-relevant objects directs attention toward them, which aligns with research showing that participants with a heightened attentional state during encoding remember images better compared to a lower attentional state (Corriveau et al., 2024; de Bettencourt et al., 2018; Wakeland-Hart et al., 2022).

4.4. Limitations

One limitation concerns the number of objects in our task. On average, 9 out of the 12 incongruently placed objects were remembered. While most objects were recalled accurately, a few responses during room recall were categorized as either semantic errors (reflecting semantic construction) or unrelated errors. Future studies should increase the number of objects experienced in the virtual apartment. Additionally, the physical layout of the virtual apartment was fixed for all participants, meaning the spatial positions of the three rooms relative to one another (e.g., bathroom on the left, kitchen on the right) were not counterbalanced. Consequently, we cannot rule out the possibility that participants partially relied on egocentric spatial coordinates – remembering an object's location as being in the room "on the left" or "straight ahead" – rather than purely encoding the room's conceptual and semantic identity (e.g., bedroom vs. kitchen). Future iterations of the paradigm may address this issue. Moreover, in this study, we tested a healthy sample with young participants; thus, generalizability of the results to a broader population, including older adults or patients with mental disorders, is limited. Finally, while our initial sample size ($n = 88$) was appropriately powered to detect a medium-sized main effect of stress, we acknowledge that the study is likely underpowered to detect smaller interaction effects. This limitation is particularly relevant for the exploratory analysis of confidence ratings, where the necessary exclusion of participants without responses in certain error bins reduced the analyzed subsample to $n = 53$. Consequently, the observed nuanced effects – such as the three-way interaction indicating increased confidence in unrelated errors among women under stress – are exploratory and must be interpreted with caution. Future studies utilizing larger sample sizes are required to reliably replicate these specific stress-by-sex interactions.

5. Conclusion

This work provides initial findings on the relationship between acute stress and semantic construction using a naturalistic and complex paradigm. Although we have not found a solution to prevent you from losing your car keys during a stressful morning, we have shown that the favored use of semantic knowledge during scenario construction is not significantly impacted by acute stress. This suggests that episodic memories remain stable even following stressful experiences prior to recognition. We propose investigating alternative time windows when exploring stress in the context of scenario construction in the future. Finally, our exploratory analysis revealed initial evidence that acute stress increased the confidence in false memories among women. This observation opens possibilities for future research to explore sex effect on confidence levels in stressful situations.

CRediT authorship contribution statement

Nicole Klein : Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Henry D. Soldan :** Writing – review & editing, Formal analysis, Data curation. **Oliver T. Wolf :** Writing – review & editing, Project administration, Funding acquisition, Conceptualization. **Christian J. Merz :** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Funding acquisition.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) within the FOR 2812: "Constructing scenarios of the past" under grant no. 397530566 (OTW) and within the SFB 1280: Extinction Learning under grant no. 316803389 – SFB1280 (projects A09 to OTW and CJM). The DFG had no role in study design, collection, analysis, interpretation of data, manuscript writing, or submitting the manuscript for publication.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors did not use generative AI technologies for preparation of this work.

Declaration of Competing Interest

All authors report no biomedical financial interest or potential conflicts of interest.

Acknowledgments

We gratefully acknowledge the help of Niklas Hampl during data collection and recruitment of participants and help in data preparation, as well as Viktoria Dochevska, Philipp Busch, Leonie Rerri, and Johanna Reiche for help during data collection. We thank Franziska Schewe for help in the preparation of the salivary cortisol analysis, Annika Mühlenkamp for the analysis of salivary cortisol concentrations, Kim Walusiacki for the analysis of alpha-amylase levels, and Dirk Moser for supervision.

Data availability

Data will be made available on request.

References

Ballou, M. R., & Sommers, M. S. (2008). Similar phenomena, different mechanisms: Semantic and phonological false memories are produced by independent mechanisms. *Memory & Cognition, 36*(8), 1450–1459. https://doi.org/10.3758/MC.36.8.1450
Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge University Press.
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
Buchanan, T. W., Tranel, D., & Adolphs, R. (2006). Impaired memory retrieval correlates with individual differences in cortisol response but not autonomic response. *Learning & Memory, 13*(3), 382–387. https://doi.org/10.1101/lm.206306
Chang, M., & Brainerd, C. J. (2021). Semantic and phonological false memory: A review of theory and data. *Journal of Memory and Language, 119*, Article 104210. https://doi.org/10.1016/j.jml.2020.104210
Cheng, S., Werning, M., & Suddendorf, T. (2016). Dissociating memory traces and scenario construction in mental time travel. *Neuroscience & Biobehavioral Reviews, 60*, 82–89. https://doi.org/10.1016/j.neubiorev.2015.11.011
Cheng, S. (2024). Distinct mechanisms and functions of episodic memory. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 379*(1913), Article 20230411. https://doi.org/10.1098/rstb.2023.0411
Coane, J. H., McBride, D. M., Huff, M. J., Chang, K., Marsh, E. M., & Smith, K. A. (2021). Manipulations of list type in the DRM paradigm: A review of how structural and conceptual similarity affect false memory. *Frontiers in Psychology, 12*, Article 668550. https://doi.org/10.3389/fpsyg.2021.668550
Corriveau, A., Chao, A., de Bettencourt, M. T., & Rosenberg, M. D. (2024). Recognition memory fluctuates with sustained attention regardless of task-relevance. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-024-02560-x
Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268
de Bettencourt, M. T., Norman, K. A., & Turk-Browne, N. B. (2018). Forgetting from lapses of sustained attention. *Psychonomic Bulletin & Review, 25*(2), 605–611. https://doi.org/10.3758/s13423-017-1309-5
Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149
Franke, G. (2000). *BSI.Brief Symptom Inventory—Deutsche Version. Manual.* Beltz.
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
Greve, A., Van Rossum, M. C. W., & Donaldson, D. I. (2007). Investigating the functional interaction between semantic and episodic memory: Convergent behavioral and electrophysiological evidence for the role of familiarity. *NeuroImage, 34*(2), 801–814. https://doi.org/10.1016/j.neuroimage.2006.07.043
Guez, J., Saar-Ashkenazy, R., Keha, E., & Tiferet-Dweck, C. (2016). The Effect of Trier Social Stress Test (TSST) on item and associative recognition of words and pictures in healthy participants. *Frontiers in Psychology, 7*, 507. https://doi.org/10.3389/fpsyg.2016.00507
Hermans, E. J., Henckens, M. J. A. G., Joëls, M., & Fernández, G. (2014). Dynamic adaptation of large-scale brain networks in response to acute stressors. *Trends in Neurosciences, 37*(6), 304–314. https://doi.org/10.1016/j.tins.2014.03.006
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
Joëls, M., & Baram, T. Z. (2009). The neuro-symphony of stress. Article 6. *Nature Reviews Neuroscience, 10*(6). https://doi.org/10.1038/nrn2632
Kirschbaum, C., Pirke, K.-M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test' - a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology, 28*, 76–81. https://doi.org/10.1159/000119004
Klein, N., Zöllner, C., Otto, T., Wolf, O. T., & Merz, C. J. (2025). Cortisol modulates hippocampus activation during semantic substitution in men. *Neurobiology of Learning and Memory, 219*, Article 108049. https://doi.org/10.1016/j.nlm.2025.108049
Klier, C., & Buratto, L. G. (2020). Stress and long-term memory retrieval: A systematic review. *Trends in Psychiatry and Psychotherapy, 42*(3), 284–291. https://doi.org/10.1590/2237-6089-2019-0077
Kudielka, B. M., & Kirschbaum, C. (2005). Sex differences in HPA axis responses to stress: A review. *Biological Psychology, 69*(1), 113–132. https://doi.org/10.1016/j.biopsycho.2004.11.009
Küppers, V., & Bayen, U. J. (2014). Inconsistency effects in source memory and compensatory schema-consistent guessing. *Quarterly Journal of Experimental Psychology, 67*(10), 2042–2059. https://doi.org/10.1080/17470218.2014.904914
Lampinen, J. M., Copeland, S. M., & Neuschatz, J. S. (2001). Recollections of things schematic: Room schemas revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(5), 1211–1222. https://doi.org/10.1037/0278-7393.27.5.1211
Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory: Fig. 1. *Learning & Memory, 12*(4), 361–366. https://doi.org/10.1101/lm.94705
Lorentz, K., Gütschow, B., & Renner, F. (1999). Evaluation of a direct alpha-amylase assay using 2-chloro-4-nitrophenyl-alpha-D-maltotrioside. *Clinical chemistry and laboratory medicine, 37*, 1053–1062. https://doi.org/10.1515/CCLM.1999.154
Macias, C., & Persaud, K. (2024). From silos to synergy: Integrating approaches to investigate the role of prior knowledge and expectations on episodic memory. *Psychonomic Bulletin & Review, 31*(6), 2390–2409. https://doi.org/10.3758/s13423-024-02505-4
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
Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64*(1), 17–24. https://doi.org/10.1111/j.2044-8295.1973.tb01322.x
Otto, T., & Rose, J. (2023). The open toolbox for behavioral research. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02199-x
Pötzl, L., Wolf, O. T., & Merz, C. J. (2023). Rapid and delayed stress effects on recognition of female and male faces. *Psychoneuroendocrinology, 150*, Article 106043. https://doi.org/10.1016/j.psyneuen.2023.106043
R Core Team (2020). 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/.
Rantala, J., Salminen, K., Isokoski, P., Nieminen, V., Karjalainen, M., Väliaho, J., Müller, P., Kontunen, A., Kallio, P., & Surakka, V. (2024). Recall of odorous objects in virtual reality. *Multimodal Technologies and Interaction, 8*(6), 42. https://doi.org/10.3390/mti8060042
Raybaut, P. (2009). *Spyder-documentation. Available Online at: Pythonhosted.Org*.
Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803–814.
RStudio Team. (2019). *R Studio* (Version R version 4.4.0 (2024-04-24 ucrt)) [Computer software]. RStudio, PBC.
Schacter, D. L. (1999). The seven sins of memory insights from psychology and cognitive neuroscience. *American Psychologist, 54*(3), 182–203.
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
Schwabe, L., Haddad, L., & Schachinger, H. (2008). HPA axis activation by a socially evaluated cold-pressor test. *Psychoneuroendocrinology, 33*(6), 890–895. https://doi.org/10.1016/j.psyneuen.2008.03.001
Schwabe, L., Joëls, M., Roozendaal, B., Wolf, O. T., & Oitzl, M. S. (2012). Stress effects on memory: An update and integration. *Neuroscience & Biobehavioral Reviews, 36*(7), 1740–1749. https://doi.org/10.1016/j.neubiorev.2011.07.002
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
Smeets, T. (2011). Acute stress impairs memory retrieval independent of time of day. *Psychoneuroendocrinology, 36*(4), 495–501. https://doi.org/10.1016/j.psyneuen.2010.08.001
Sommet, N., & Morselli, D. (2017). Keep calm and learn multilevel logistic modeling: A simplified three-step procedure using Stata, R, Mplus, and SPSS. Article 1. *International Review of Social Psychology, 30*(1). https://doi.org/10.5334/irsp.90
Staresina, B. P., Gray, J. C., & Davachi, L. (2009). Event congruency enhances episodic memory encoding through semantic elaboration and relational binding. *Cerebral Cortex, 19*(5), 1198–1207. https://doi.org/10.1093/cercor/bhn165
The MathWorks Inc. (2021). MATLAB (Version MATLAB version: R2021a). https://www.mathworks.com.
Tulving, E. (1972). Episodic and semantic memory. *Organization of Memory, 1*, 381–403.
Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53*, 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114
Ulrich-Lai, Y. M., & Herman, J. P. (2009). Neural regulation of endocrine and autonomic stress responses. Article 6. *Nature Reviews Neuroscience, 10*(6). https://doi.org/10.1038/nrn2647
Umanath, S., & Coane, J. H. (2020). Face Validity of Remembering and Knowing: Empirical Consensus and Disagreement Between Participants and Researchers. *Perspectives on Psychological Science, 15*(6), 1400–1422. https://doi.org/10.1177/1745691620917672
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 Jr, F. L. (2009). *Python 3 Reference Manual: (Python Documentation Manual Part 2)*. Createspace.
Wakeland-Hart, C. D., Cao, S. A., de Bettencourt, M. T., Bainbridge, W. A., & Rosenberg, M. D. (2022). Predicting visual memory across images and within individuals. *Cognition, 227*, Article 105201. https://doi.org/10.1016/j.cognition.2022.105201
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
Yonelinas, A. P. (2001). Components of episodic memory: The contribution of recollection and familiarity. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences, 356*(1413), 1363–1374. https://doi.org/10.1098/rstb.2001.0939
Yonelinas, A. P., Ranganath, C., Ekstrom, A. D., & Wiltgen, B. J. (2019). A contextual binding theory of episodic memory: Systems consolidation reconsidered. *Nature Reviews Neuroscience, 20*(6), 364–375. https://doi.org/10.1038/s41583-019-0150-4
Zaragoza, M. S., & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 934–945.
Zöllner, C., Klein, N., Cheng, S., Schubotz, R. I., Axmacher, N., & Wolf, O. T. (2023). Where was the toaster? A systematic investigation of semantic construction in a new virtual episodic memory paradigm. *Quarterly Journal of Experimental Psychology, 174702182211166*. https://doi.org/10.1177/17470218221116610