Article

Current Biology

# The memory trace of a stressful episode

## Highlights

- Psychosocial stress shapes representations of memory traces in the human amygdala
- Neural representations of central items of a stressful episode are bound together
- Representations of remembered central items bind to representations of the stressor
- Amygdala similarity patterns explain how stress improves memory for central items

## Authors

Anne Bierbrauer, Marie-Christin Fellner, Rebekka Heinen, Oliver T. Wolf, Nikolai Axmacher

## Correspondence

anne.bierbrauer@rub.de (A.B.), oliver.t.wolf@rub.de (O.T.W.), nikolai.axmacher@rub.de (N.A.)

## In brief

Bierbrauer et al. identify the neural representations of specific objects from a stressful episode in the human amygdala. They show that stress leads to neural generalization (higher representational similarity) among central objects of the episode, and that the similarity of these objects to the stressor can explain why they are better remembered.

Bierbrauer et al., 2021, Current Biology *31*, 1–10
December 6, 2021 © 2021 Ruhr University Bochum. Published by Elsevier Inc.
https://doi.org/10.1016/j.cub.2021.09.044

CellPress

Please cite this article in press as: Bierbrauer et al., The memory trace of a stressful episode, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.09.044

Current Biology

CellPress
OPEN ACCESS

Article

# The memory trace of a stressful episode

Anne Bierbrauer,[1,5,*] Marie-Christin Fellner,[1] Rebekka Heinen,[1] Oliver T. Wolf,[2,4,*] and Nikolai Axmacher[1,3,4,*]
[1]Department of Neuropsychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr University Bochum, Universitätsstraße 150, 44801 Bochum, Germany
[2]Department of Cognitive Psychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr University Bochum, Universitätsstraße 150, 44801 Bochum, Germany
[3]State Key Laboratory of Cognitive Neuroscience and Learning and IDG/McGovern Institute for Brain Research, Beijing Normal University, Xinjiekouwai Street 19, Beijing 100875, China
[4]These authors contributed equally
[5]Lead contact
*Correspondence: anne.bierbrauer@rub.de (A.B.), oliver.t.wolf@rub.de (O.T.W.), nikolai.axmacher@rub.de (N.A.)
https://doi.org/10.1016/j.cub.2021.09.044

## SUMMARY

Stress influences episodic memory formation via noradrenaline and glucocorticoid effects on amygdala and hippocampus. A common finding is the improvement of memory for central aspects of a stressful episode. This is putatively related to changes in the neural representations of specific experiences, i.e., their memory traces. Here we show that the memory improvement for objects that were encountered in a stressful episode relates to differences in the neural representations of these objects in the amygdala. Using functional magnetic resonance imaging, we found that stress specifically altered the representations of central objects: compared to control objects, they became more similar to one another and more distinct from objects that were not part of this episode. Furthermore, higher similarity of central objects to the main stressor—the faces of the stress-inducing committee members—predicted better memory. This suggests that the central objects were closely integrated into a stressor-centered memory representation. Our findings provide mechanistic insights into how stress shapes the memory trace and have profound implications for neurocognitive models of stressful and emotional memory.

## INTRODUCTION

Psychosocial stress exerts profound effects on episodic memory encoding and consolidation.[1,2] Among others, stress induces the release of noradrenaline and glucocorticoids, which influence neural processing in amygdala, hippocampus, and various other brain regions.[3] Effects on activity in the amygdala seem to be particularly important for memories of stressful and emotional events.[3,4] While stress during encoding may also be detrimental,[5] several studies have shown that specifically memory for central aspects of a stressful episode is enhanced, suggesting that stress exerts distinct influences on some memory contents and leaves others relatively unaffected.[5–7] However, how stress influences the neural representations of specific events has remained elusive, in particular in the human brain.

Representational similarity analysis (RSA) is a well-established method to investigate the structure of neural representations.[8] More specifically, RSA can be employed to characterize neural population codes. Representational similarity (or its reverse, representational distance) can be quantified using different measures such as Euclidean distance, Mahalanobis distance, or Pearson/Spearman correlations. Representational similarity matrices reflect the similarities between voxel patterns of individual stimuli and describe the "geometry" of neural representations[9] (Figure 1B). RSA is fundamentally different from classic univariate approaches as the condition differences do not relate to overall activity differences, but to representational relationships between the distributed response patterns in a given brain region. Previous studies employed RSA to characterize how individual items are represented in human episodic memory[10,11] and showed how they are modulated by various cognitive and emotional factors that influence memory.[11,12]

Here we used RSA to uncover the representational structure of memory traces for a real-life stressful episode. The Trier social stress test (TSST) was employed as a well-established and ecologically valid paradigm to induce psychosocial stress.[13] TSST effects were compared to those of a non-stressful control episode, the "friendly" TSST[14] (f-TSST; Figure 1A, left). In the TSST, participants conduct mock job interviews in front of a two-person evaluation committee acting in a neutral and reserved manner. In the control f-TSST, they can freely choose to talk about their career aspirations and hobbies, and the committee members react in an encouraging and friendly manner.[14]

Critically, participants incidentally encountered 24 objects in either episode (e.g., a cola can or a teapot; see Figure S1 for pictures of all objects). The committee members utilized half of these objects, e.g., taking a sip of tea, thereby making them "central" to the episode[6] (Figure S1). On the following day, we measured episodic memory performance of all objects and used functional magnetic resonance imaging (fMRI) to assess the neural representations of the objects and of the stressors, i.e., the faces of the committee members (Figure 1A, right). In


This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Please cite this article in press as: Bierbrauer et al., The memory trace of a stressful episode, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.09.044



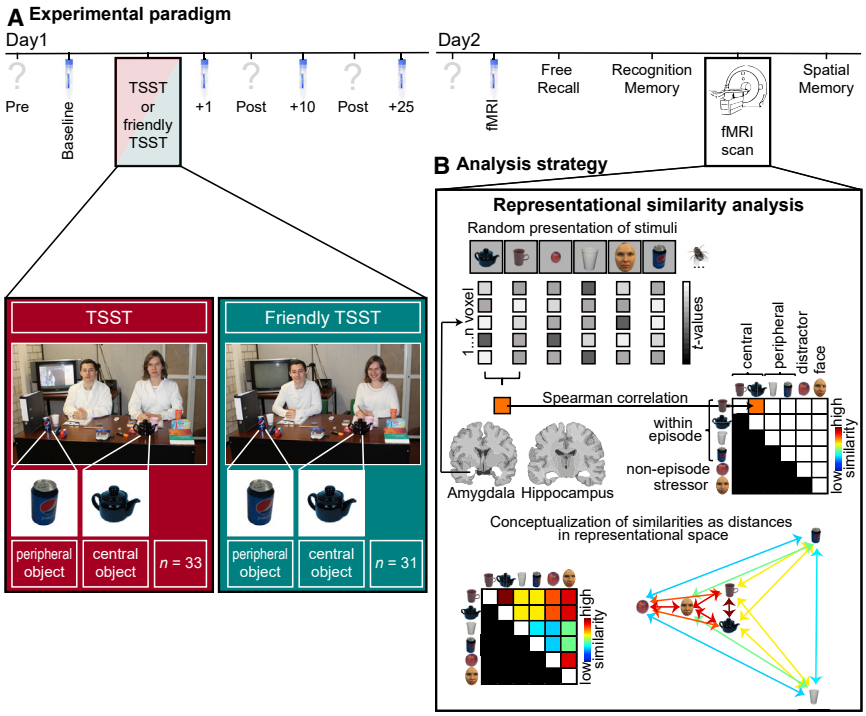


**Figure 1. Experimental paradigm and analysis strategy**

(A) The experiment was performed on two consecutive days. On day 1, participants conducted either the Trier social stress test (TSST) or a non-stressful control test ("friendly TSST," f-TSST). They provided saliva samples to assess cortisol levels and rated their affect and self-esteem before and after the (f-)TSST. On day 2, they performed a free recall and a recognition memory test of objects that occurred in the (f-)TSST and were then presented pictures of these objects and of the committee members in the MRI scanner.

(B) Analysis strategy. Top: we applied representational similarity analysis on data from the amygdala and the hippocampus by correlating voxel patterns between pairs of pictures. More specifically, we extracted the similarity between voxel patterns as an indicator of how similar two objects, or an object and a face, were represented in the brain (e.g., LaRocque et al.[10]). Bottom: higher correlations indicate that voxel patterns are more similar to one another, i.e., that their distance in representational space is lower.

See also Figure S1 and Table S1.

the recognition memory test as well as during the fMRI recordings, we presented pictures of the 24 objects from the (f-)TSST session together with 24 "difficult" distractors (i.e., objects that were similar to those that had actually been in the session) and 24 "easy" distractors (i.e., objects that have not been in the session). Moreover, we presented pictures of the committee members' faces and of similar distractor faces.

Previous behavioral studies showed a specific memory benefit for central over peripheral objects in the stressful condition.[6,15] However, different conceptual frameworks made opposing predictions on how stress and emotions alter the representations of specific events in episodic memory: on one hand, stress may increase the distinctiveness of the neural representations of central objects. This could be related to pattern separation processes, a key mechanism for episodic memory.[10,16,17] Higher distinctiveness of individual object representations may thus improve memory by disentangling them from one another and from the representations of new, distractor objects.[10,11,18,19] On the other hand, stress—or more generally, negative emotional valence—may enhance memory by binding all relevant aspects of a stressful episode together.[7,20] The representations of individual features from these episodes may generalize and become more similar to the representation of the stressor and of the associated negative emotion ("emotional binding"[20]). In contrast to the idea that memory is supported by higher distinctiveness of individual item representations, memory would be enhanced by integrating all aspects of a stressful episode into one, tightly bound, coherent representation. More specifically, this may benefit episodic memory as representations could be fully reactivated by various associated retrieval cues.

We particularly focused on neural representations in hippocampus and amygdala. The hippocampus has been shown to play a critical role in supporting both distinct and bound representations of episodic memories.[10,11,16] Hippocampal neurons in both humans and rodents display pronounced mixed selectivity:[21] they respond to various features of an event, depending on task demands.[22,23] This is consistent with the idea that the hippocampus serves as an "index" to multimodal and multifaceted representations in the neocortex.[24]

Engrams of fear memories have been identified in rodent hippocampus and amygdala.[25,26] Recent evidence has further shown that amygdala neurons do not only respond to fearful or stress-related stimuli, but exhibit mixed selectivity as well:[21,27] their firing may represent various different emotional and social dimensions, depending on task and context.[28] In humans, amygdala neurons respond to faces and to perceived emotions,[29,30] and fMRI studies showed that the amygdala represents both fear memories and the subjective valence of odors.[12,31,32] Such multidimensional representations may serve to bind the diverse aspects of an emotional experience into one integrated episode.

We analyzed the representational similarity between central and peripheral objects from the TSST and the f-TSST. Our results speak in favor of the "emotional binding" hypothesis: compared to peripheral objects and to objects in the non-stressful control condition, amygdala representations of central TSST objects became more similar to one another and less similar to control objects that were not part of the episode. Moreover, their representational similarity with the stressors predicted later memory, accounting for the superior memory for central events from a stressful episode.

Please cite this article in press as: Bierbrauer et al., The memory trace of a stressful episode, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.09.044



CellPress
OPEN ACCESS

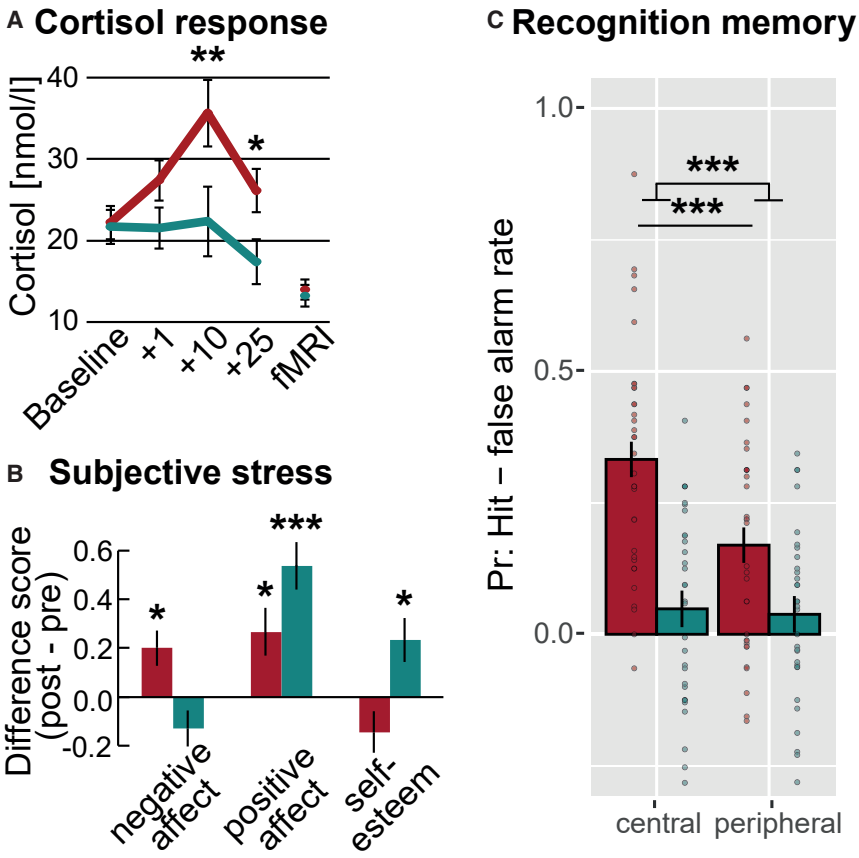


**Figure 2. TSST induces a cortisol response, negative affect, and better recognition memory performance for central objects**

(A) Cortisol increased significantly in the TSST group (red), but not in the f-TSST group (cyan).

(B) Differences in negative affect, positive affect, and state self-esteem after versus before the experimental intervention.

(C) Recognition memory performance Pr (hit – false alarm rate) was higher for central compared to peripheral objects in the stressful condition only.

Dots show descriptive means for each participant; error bars show standard errors; $^{*}p_{corrected} < 0.05$; $^{**}p_{corrected} < 0.01$; $^{***}p_{corrected} < 0.001$; see also Figure S2 and Table S3.

## RESULTS

We compared two groups of participants (Figure 1A, left): one group undergoing the TSST (n = 33), another one the f-TSST (n = 31). The groups did not differ in terms of age, trait anxiety, self-esteem, depression score, wake-up and testing time on day 1, and baseline scores of affect and cortisol on day 1 or day 2 (Table S1). As expected, the TSST led to a significant physiological and subjective stress response: the concentration of saliva cortisol increased, while affect and self-esteem were negatively impacted compared to the f-TSST (Figures 2A and 2B; STAR Methods).

### Stress boosts memory performance for salient components of the episode

We analyzed three measures of memory performance: the number of freely recalled objects, recognition memory for these objects (Pr = hit – false alarm rate), and spatial memory of object positions in the (f-)TSST. We applied a mixed ANOVA with "object type" (central versus peripheral) as within-subject factor and "experimental condition" (TSST versus f-TSST) as between-subject factor. In the following paragraph, we describe only effects on recognition memory; effects on free recall and spatial memory were similar (Figure S2; Table S3).

Recognition memory performance was higher for central versus peripheral objects ($F(1,62) = 7.73$, $p = 0.007$; $f^2 = 0.12$; Figure 2C) and for TSST versus f-TSST participants ($F(1,62) = 31.49$, $p < 0.001$; $f^2 = 0.51$). A significant interaction ($F(1,62) = 5.99$, $p = 0.017$; $f^2 = 0.10$) indicated more pronounced stress effects on central than peripheral objects, replicating previous findings.[1,6,15] Simple effects analyses showed better memory in the TSST than the f-TSST for central objects ($t(120.5) = 5.87$, $p_{Bonferroni} < 0.001$; $f^2 = 0.29$) and for peripheral objects ($t(120.5) = 2.72$, $p_{Bonferroni} = 0.015$; $f^2 = 0.06$). In the TSST, central objects were better remembered than peripheral objects ($t(62) = 3.76$, $p_{Bonferroni} < 0.001$; $f^2 = 0.23$). This was not the case in the non-stressful control condition ($t(62) = 0.23$, $p_{Bonferroni} \approx 1$).

In sum, we found that stress generally improved memory performance, but this effect was specifically pronounced for central objects, replicating previous results.[1,6,15] Control analyses showed that effects relied on better memory for old objects (i.e., hits) rather than fewer false alarms (to either difficult or easy distractors; Figure S2; Table S3).

### Stress alters the representational structure of the memory trace

Stress effects on memory formation are likely mediated by the influence of noradrenaline and glucocorticoids on neural processing in amygdala and hippocampus.[3,33,34] As previous functional imaging studies point to a possible lateralization of stress effects in these regions,[35,36] we focused our analysis on neural representations in four regions of interest (ROIs), i.e., left/right hippocampus and left/right amygdala.

We analyzed the representational similarity of voxel patterns (1) among central and among peripheral objects, (2) between these objects and distractors, and (3) between these objects and the "stressors," i.e., the faces of the committee members (Figure 1B). All outcome variables were analyzed at the level of individual objects in a linear mixed model with "object type" (central versus peripheral) as within-subject factor, "experimental condition" (TSST versus f-TSST) as between-subject factor, and "subject" as random factor. "Representational reliability" (i.e., the similarity of representations of the same object across trials; STAR Methods) was included as a covariate (Table S2).

Please cite this article in press as: Bierbrauer et al., The memory trace of a stressful episode, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.09.044



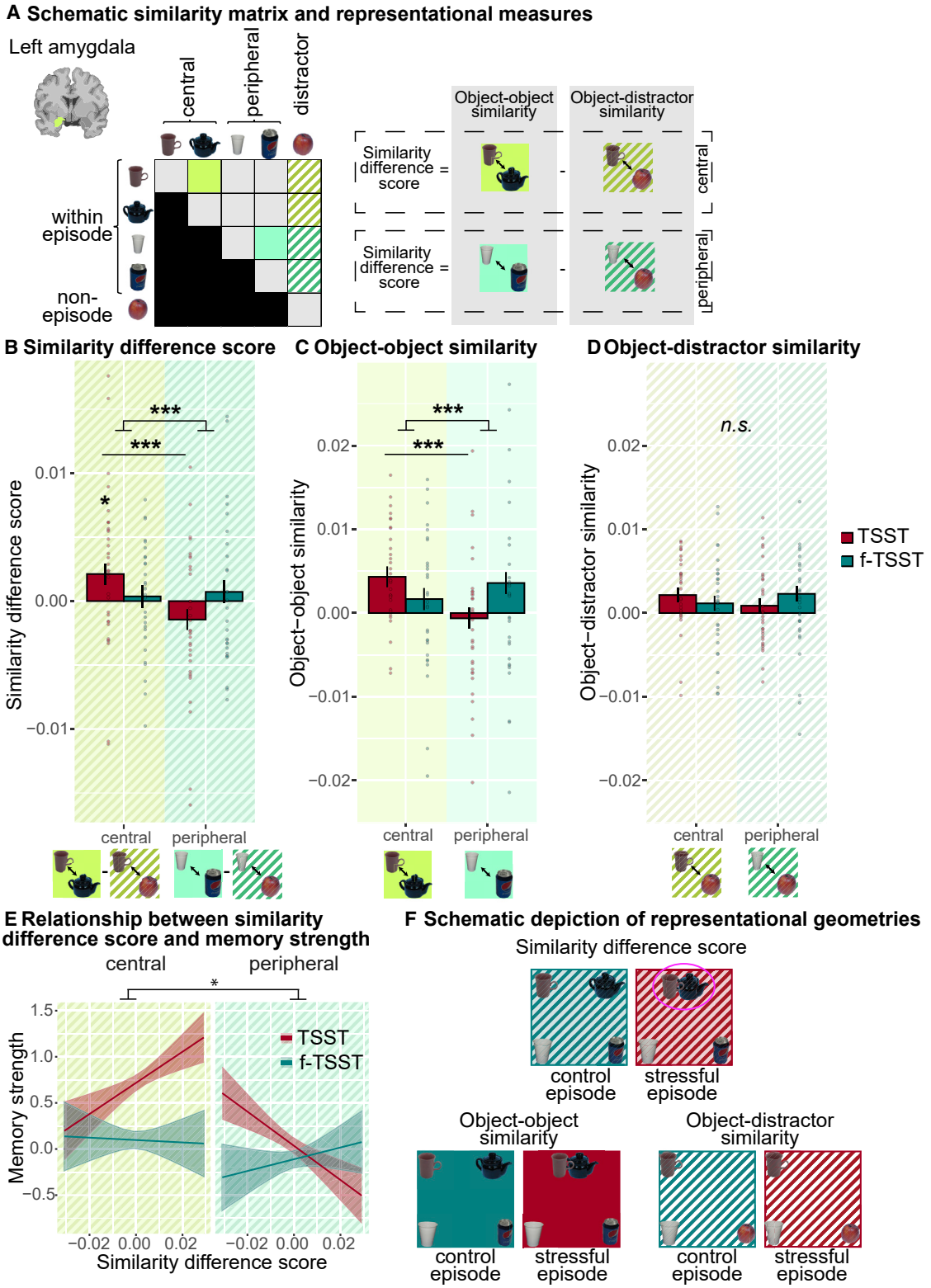


**Figure 3. Stress promotes a higher similarity difference score and induces a generalization of representations of central objects in the left amygdala**

(A) Schematic depiction of representational similarity analysis in the left amygdala. We compared the neural similarities among central and among peripheral objects in the TSST and the f-TSST ("object-object similarity"), and between these objects and distractor objects ("object-distractor similarity"). "Similarity

*(legend continued on next page)*

Please cite this article in press as: Bierbrauer et al., The memory trace of a stressful episode, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.09.044



CellPress
OPEN ACCESS

### Episodic binding is higher for central objects of a stressful episode in the left amygdala

Remembering objects from the TSST or the f-TSST requires participants to distinguish between objects that were part of the episode (central or peripheral objects) versus those that were not (distractor objects). We thus calculated the difference in similarities among objects that were part of the episode ("object-object similarity") and contrasted it with the similarity between objects from the episode and distractors ("object-distractor similarity"). This difference score indicates the degree to which objects from the episode were more similar to each other than to distractors that were not part of the episode ("similarity difference score"; Figure 3A). We compared the similarity difference score between object types (central versus peripheral) and between experimental conditions (TSST versus f-TSST).

The similarity difference score was higher for central than for peripheral objects (main effect of object type: $F(1,954.8) = 8.81$, $p_{Bonferroni} = 0.012$; $f^2 = 0.01$). There was no main effect of stress ($F(1,54) = 0.04$, $p_{Bonferroni} \approx 1$), but a significant interaction ($F(1,954.7) = 13.35$, $p_{Bonferroni} = 0.001$; $f^2 = 0.01$; Figure 3B). Follow-up tests showed that in the stressful condition, central objects showed a higher similarity difference score compared to peripheral objects ($t(954.8) = 4.85$, $p_{Bonferroni} < 0.001$; $f^2 = 0.02$). This effect was not found in the f-TSST ($t(954.7) = -0.45$, $p_{Bonferroni} \approx 1$). Only the similarity difference score of central objects in the stressful condition differed significantly from zero ($t(84.2) = 2.59$, $p_{Bonferroni} = 0.023$; $f^2 = 0.08$; f-TSST: $t(84.5) = -0.42$, $p_{Bonferroni} \approx 1$; peripheral objects: both $t \leq 1.76$, both $p_{Bonferroni} \geq 0.163$). Thus, the representations of these objects were more similar to each other than they were to distractor objects, and thus moved closer in the representational space of the left amygdala (Figure 3F, upper part).

Analyzing easy and difficult distractors separately, we found that this interaction effect was restricted to easy distractors ($F(1,954.0) = 11.91$, $p_{Bonferroni} < 0.001$; $f^2 = 0.01$; difficult distractors: $F(1,950.0) = 0.02$, $p_{Bonferroni} = 0.892$): again, similarity difference scores were higher for central versus peripheral objects in the TSST ($t(954.0) = 4.66$, $p_{Bonferroni} < 0.001$; $f^2 = 0.02$), but not in the f-TSST condition ($t(954.0) = -0.35$, $p_{Bonferroni} \approx 1$). We did not find any effects in the other ROIs (right amygdala or left and right hippocampus: all $F \leq 4.37$, all $p_{Bonferroni} \geq 0.147$; Figure S3; Table S4).

### Stress promotes generalized representations of central objects in the left amygdala

As described above, the similarity difference score compares the similarity among objects that were part of the episode ("object-object similarity") to the similarity between these objects and distractor objects that were not part of the episode ("object-distractor similarity"). The effect may thus result from (1) higher object-object similarity, (2) lower object-distractor similarity, or (3) both. In order to better understand the factors contributing to the effects on the similarity difference score, we analyzed its two components separately, i.e., object-object similarity and object-distractor similarity (Figure 3A). Higher generalization indicates that representations become more similar, i.e., more adjacent in hippocampus or amygdala representational space; lower generalization means that representations are more distinct, i.e., more remote in hippocampus or amygdala representational space (Figure 3F).

We first analyzed object-object similarity in the left amygdala as a function of object type (central versus peripheral) and stress (TSST versus f-TSST). There were no main effects of object type ($F(1,969.7) = 5.10$, $p_{Bonferroni} = 0.097$) or stress ($F(1,55) = 0.21$, $p_{Bonferroni} \approx 1$), but a significant interaction ($F(1,969.7) = 26.42$, $p_{Bonferroni} < 0.001$; $f^2 = 0.03$; Figure 3C). Follow-up analyses showed that representations of central objects generalized more than representations of peripheral objects following the TSST ($t(969.7) = 5.36$, $p_{Bonferroni} < 0.001$; $f^2 = 0.03$), but not after the f-TSST ($t(969.7) = -1.97$, $p_{Bonferroni} = 0.099$; Figure 3F, lower left part). Interestingly, the amount of generalization of peripheral objects in the f-TSST condition was similar to the generalization of central objects in the TSST condition. This result pattern suggests that the stress effect is driven by both a reduction in similarity among peripheral objects and an increase in similarity among central objects. We did not find any effects in the other ROIs (all $F \leq 3.57$, all $p_{Bonferroni} \geq 0.237$; Figure S3; Table S4).

We next analyzed possible object type or stress effects on the similarity between objects from the (f-)TSST and distractor objects (i.e., object-distractor similarity). However, we did not observe any effect of stress ($F(1,55) = 0.04$, $p_{Bonferroni} \approx 1$) or object type ($F(1,971) = 0.02$, $p_{Bonferroni} \approx 1$) and no interaction ($F(1,970.9) = 5.39$, $p_{Bonferroni} = 0.082$; Figure 3D). Object-distractor similarity did not differ between central and peripheral objects in the TSST ($t(970.9) = 1.77$, $p_{Bonferroni} = 0.154$; Figure 3F, lower right part). We also did not find any effects in the other ROIs (all $F \leq 1.26$, all $p_{Bonferroni} \approx 1$; Figure S3; Table S4).

Taken together, we found that stress selectively increases the generalization of neural representations of central compared to peripheral objects in the left amygdala. We did not observe such an effect for the similarity of these objects to distractors that were not part of the stressful episode. Stress selectively moved the neural representations of central objects closer together in representational space and not further apart, as would have been predicted by the distinctiveness hypothesis (Figure 3F). This may improve episodic memory as objects that were part of the episode are tightly bound together and thereby separated from distractor objects.

difference score" refers to the difference between object-object similarity and object-distractor similarity and quantifies the binding of objects within the (f-)TSST episode.

(B) The similarity difference score was higher for central versus peripheral objects in the TSST (red bars), but not in the f-TSST (cyan bars).

(C) Higher object-object similarity among central versus among peripheral objects in the TSST, but not in the f-TSST.

(D) Object-distractor similarity did not differ between object types or between TSST and f-TSST.

(E) Significant three-way interaction: the relationship between the similarity difference score and memory strength differed between conditions. However, the similarity difference score could not predict memory strength for any of the object types or experimental conditions.

(F) In a stressful episode, central objects moved closer in representational space.

Dots show descriptive means for each participant; error bars show standard errors; n.s., not significant; $^{*}p_{corrected} < 0.05$; $^{***}p_{corrected} < 0.001$; see also Figures S3 and S4 and Tables S2 and S4.

Please cite this article in press as: Bierbrauer et al., The memory trace of a stressful episode, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.09.044



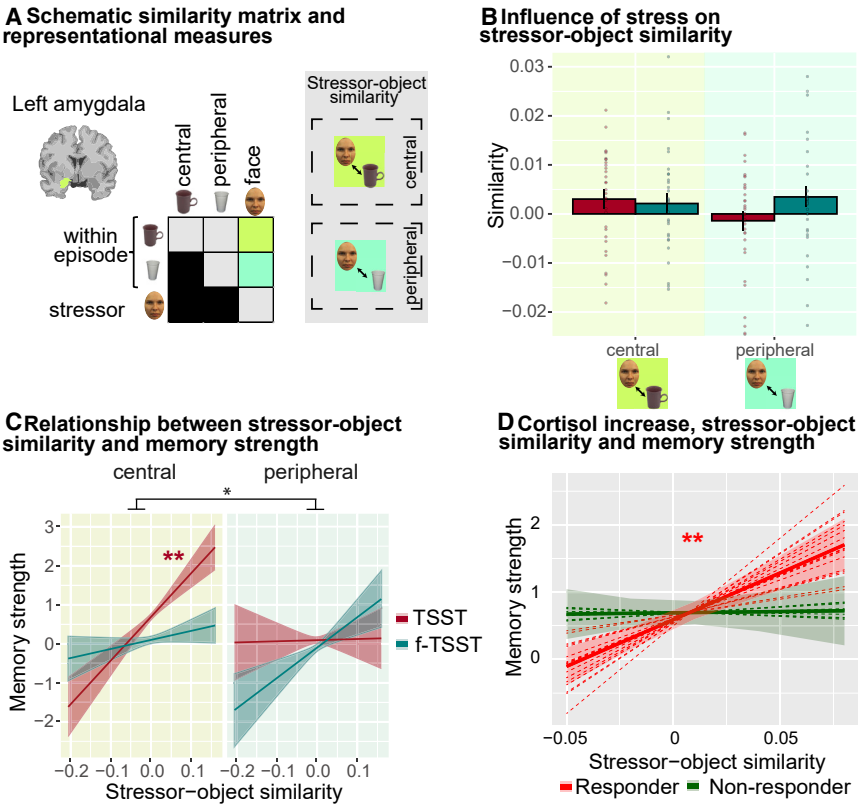


**Figure 4. Stressor-object similarity accounts for memory benefits of stress**

(A) Schematic depiction of representational similarity analysis in the left amygdala. Stressor-object similarity refers to the representational similarity of central or peripheral objects to the faces of the committee members.

(B) Stressor-object similarity does not differ between object types or between TSST (red bars) and f-TSST (cyan bars).

(C) Significant three-way interaction: stressor-object similarity predicts memory strength for central objects in the TSST (red lines), but not in the f-TSST (cyan lines).

(D) Selective effects of stressor-object similarity on memory strength for cortisol responders; depicted are model estimates for responders (red) and non-responders (green) and for single participants in the respective groups (dashed lines).

(E) Central objects that moved closer to the stressor in representational space were more likely to be remembered.

Dots show descriptive means for each participant; error bars show standard errors; $^{*}p_{corrected} < 0.05$; $^{**}p_{corrected} < 0.01$; see also Figures S3 and S4 and Tables S2 and S4.

### *Stress effects on the similarity difference score cannot predict memory benefits of central objects*

We next tested whether stress effects on the similarity difference score may explain the selective memory enhancement of central objects in the stressful episode. We calculated the memory strength of each object as the rating of this object in the recognition memory test from which we subtracted the average rating of all objects by a given participant. These values were then inverted, such that higher values reflected higher memory strength (for details, see STAR Methods). We created a mixed linear model to predict recognition memory strength at the level of single objects using "object type" (central versus peripheral) and "similarity difference score" as within-subject factors, "experimental condition" (TSST versus f-TSST) as between-subject factor, "subject" as random factor, and "representational reliability" as covariate.

We observed a significant three-way interaction between stress, object type, and similarity difference score ($F(1,782.2) = 4.26$, $p = 0.039$; $f^2 = 0.01$; Figure 3E). However, when we separately analyzed the effect of the similarity difference score on memory for the different object types and experimental conditions, these relationships were not significant (tests for slopes against 0: all $t \leq 1.99$, all $p_{Bonferroni} \geq 0.094$). Thus, while stress increases the similarity difference score of central objects, this effect cannot explain the stress-induced memory benefit of central objects.

### *Stressor-object similarity predicts the similarity difference score and memory benefits of stress*

Next, we asked whether the stress-induced binding of central objects may be related to their representational relationship to the stressor, and whether the latter predicts memory performance. There are various aspects in the TSST that cause stress, but arguably the strongest consists of the committee.[37] We therefore hypothesized that the memory benefit for central objects in the stressful condition can be predicted by the similarity of the object representations to the representation of the main stressor of the experiment, the faces of the committee members (Figure 4A).

We assigned each central object to the committee member who had manipulated this object: e.g., the female committee member took the mug to prepare tea, while the male committee member used the ruler to underline a word on his clipboard (Figure S1). We calculated the similarity between the neural representations of objects and stressor ("stressor-object similarity"). For peripheral objects, we used the similarity of the objects to both committee members. We again tested for differences between object types and experimental conditions.

This analysis did not show any main effects of either stress or object type on stressor-object similarity in the left amygdala (both $F \leq 1.01$, both $p \geq 0.314$), but a trend for an interaction

Please cite this article in press as: Bierbrauer et al., The memory trace of a stressful episode, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.09.044



CellPress
OPEN ACCESS

($F(940.3) = 3.62$, $p = 0.057$; Figure 4B). The pattern of results was descriptively similar to those of the similarity difference score and showed a trend for higher stressor-object similarity of central compared to peripheral objects in TSST participants ($t(940.4) = 2.11$, $p_{Bonferroni} = 0.070$; Figure 4E, left part).

In an additional analysis, we found that stressor-object similarity predicted the similarity difference score ($F(1,896.4) = 8.01$, $p = 0.005$; $f^2 = 0.01$), suggesting that the central objects are bound within the episode because their representations become more similar to the representation of the stressor.

We next analyzed whether stressor-object similarity in the left amygdala may predict memory performance. Indeed, we found that higher stressor-object similarity predicted better memory ($F(1,937.7) = 7.35$, $p = 0.007$; $f^2 = 0.01$). In addition, we observed a significant three-way interaction between stressor-object similarity, object type, and experimental condition ($F(1,980.9) = 4.24$, $p = 0.040$; $f^2 = 0.004$; Figure 4C). Follow-up analyses revealed a highly specific effect, whereby stressor-object similarity predicted memory only for central objects in the TSST condition ($t(964) = 3.08$, $p_{Bonferroni} = 0.004$; $f^2 = 0.01$), but not in any of the other conditions (all $t \leq 1.74$, all $p_{Bonferroni} \geq 0.166$). In other words, those central objects in the stressful episode whose left amygdala representations became more similar to the representations of the stress-inducing faces were better remembered (Figure 4E, right part).

In a model including both stressor-object similarity and the similarity difference score, we found that stressor-object similarity, but not the similarity difference score, predicted memory for the central objects in the TSST condition (main effect of stressor-object similarity: $F(1,895.9) = 4.84$, $p = 0.028$; $f^2 = 0.01$; three-way interaction with object type and experimental condition: $F(1,940.6) = 7.18$, $p = 0.007$; $f^2 = 0.01$; slope for predicting memory performance of central objects in the TSST condition: $t(924) = 3.05$, $p_{Bonferroni} = 0.005$; $f^2 = 0.01$; similarity difference score: all $p \geq 0.072$).

Further analyses showed that this effect was restricted to cortisol responders, i.e., participants in the TSST condition with a saliva cortisol increase of more than 1.5 nmol/L.[38] We directly compared the relationship between stressor-object similarity and memory for cortisol responders and cortisol non-responders: cortisol responders ($n = 19$) showed a significant relationship between stressor-object similarity and memory strength ($t(235) = 2.93$, $p_{Bonferroni} = 0.008$; $f^2 = 0.04$; non-responders: $n = 9$; $t(238) = 0.02$, $p_{Bonferroni} \approx 1$; Figure 4D). Higher proximity of central objects and the stressor in the representational space of the left amygdala was predictive of better memory (Figure 4E, right part).

Finally, we performed two additional analyses to understand the effects in detail: (1) we added the representations in the right amygdala and a predictor for "hemisphere" (left versus right). We found that stress effects on neural representations were indeed specific to the left amygdala (Figure S4; Table S4). (2) We found similar effects of stress on neural representations when restricting the analysis to remembered objects (Figure S4; Table S4).

## DISCUSSION

Why do we remember central aspects of a stressful episode particularly well? This fundamental question is relevant for understanding both everyday memories and memory distortions in mental disorders.[1] In the current study, we addressed this question using a design with high ecological validity because participants incidentally encoded objects in an episode that closely resembles stressful real-life experiences.

We showed that stress systematically modulates the representations of objects from the stressful episode in the amygdala: stress integrates central objects from the episode and binds them to the main stressor. These findings speak against the notion that stress improves memory by increasing the distinctiveness of individual object representations. Instead, they are in line with concepts of "emotional binding,"[7,20] which suggest that the amygdala integrates items and emotions within emotional episodes—creating memory traces with tightly bound and highly memorable representational structures. Binding the various aspects that we encounter in an episode is key for episodic memory.[39,40] While binding of neutral items and contexts relies on the hippocampus,[41,42] emotional binding seems to depend on the amygdala.[7,20] Different theoretical frameworks have elaborated on the role of emotional arousal—a central part of stress reactions—for amygdala-dependent memory formation.

Mather highlighted the relevance of the amygdala for binding due to arousal.[7] According to her object-based framework, emotionally arousing objects attract attention that enhances binding of constituent features. Consequently, an object that elicits strong emotions is better remembered than a neutral object. We previously suggested that arousal does not have to be an inherent feature of an item, i.e., an item does not have to be arousing itself, but a stressful situation can render central objects more memorable.[6] Our new findings support and extend these ideas: we show that it is indeed the stressor, i.e., a core feature of the stressful situation, whose representation predicts the memorability of central items (which are inherently neutral). More specifically, we found that stress increases binding among the representations of central items, and that this effect depends on increased representational proximity of objects to the stressor, which in turn goes along with better memory.

A stressful episode is not equivalent to an emotional episode. However, acute psychosocial stress induces profound negative emotions, including shame and anxiety.[43] We observed pronounced increases of negative affect after the TSST. It seems thus evident to apply emotional binding theories to memories of a stressful episode. Yonelinas and Ritchey[20] proposed that emotional episodes activate the amygdala, which ties the emotional response to representations of individual items. Consistent with this idea and in line with various studies emphasizing the role of the amygdala for emotional memories,[44,45] we found that stress increased the binding among representations of central objects in the left amygdala.

From a different perspective, these effects may also be interpreted as a higher-order distinctiveness of objects from one episode to distractor objects: the higher similarity difference score of central objects in stressed participants implies that the stressful episode as a whole becomes more distinct from other potential episodes. This suggests that the concepts of generalization and distinctiveness are not necessarily mutually exclusive, but may occur on different levels and conjointly improve memory. Interestingly, LaRocque et al.[10] showed that higher similarities among items from the same category in

Please cite this article in press as: Bierbrauer et al., The memory trace of a stressful episode, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.09.044



perirhinal and parahippocampal cortex were beneficial for their memory. One may speculate that different brain regions promote memory by increasing representational similarities along different dimensions, i.e., objects in perirhinal cortex, scenes in parahippocampal cortex, and stress-related objects in the amygdala.

Previous studies indicated that stress has the strongest benefit on gist-like memory rather than detailed memory.[46,47] This would have predicted higher false alarm rates for difficult distractors, an effect that we did not find. On the other hand, our fMRI results suggest that the easy rather than the difficult distractors drive the effect of experimental condition and object type on the similarity difference score. The amygdala thus represents central TSST objects only distinct from easy but not difficult distractors, consistent with a gist-like representation.

Our finding of representational changes in the amygdala does not rule out that neural representations in the right amygdala or other brain regions are involved in the episodic memory of a stressful episode. However, direct comparisons between left and right hemisphere indicate that both the representational geometries and the prediction of memory strength by stressor-object similarity are indeed specific to the left amygdala. These results are consistent with meta-analyses of fMRI studies on emotional processing that showed a larger number of activation peaks in the left as compared to the right amygdala.[36,48] This may relate to more sustained temporal dynamics of the left amygdala in response to emotional stimuli.[48–51] Moreover, we also expected the hippocampus to be implicated in the memory representation but did not observe similar effects as in the amygdala. It is possible that an explicit retrieval task inside the scanner would have revealed effects on hippocampal representations.

In fact, we cannot unequivocally exclude that generalization and binding effects could have been (partly) produced during retrieval and not during encoding or consolidation. It is likely that the retrieval has changed the neural representations of the objects. Future studies may tackle important questions about the specificity of our effects with respect to the memory process they concern and whether they are indeed specific to stress or rather relate to better memory more generally.

## Conclusions

Our study demonstrates how a real-life stressful episode is represented in episodic memory and differs from the representation of a neutral episode. We found that stress affected neural representations in the amygdala: objects that were central to the stressful episode because they were manipulated by the experimenters had similar representations in the amygdala. Intriguingly, their representations also became similar to representations of the stressor's face, and this effect was functionally relevant for memory. Our findings address key predictions from prominent theories on emotional and stressful memories and may serve as a novel framework for translational research into the psychopathology of memory.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:



## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.09.044.

## ACKNOWLEDGMENTS

We are grateful to all participants who volunteered to participate in this study. We thank Uta S. Wiemers and Nadja Herten for support with the implementation of the paradigm. We thank Alex Babicz, Andreas Haltermann, Thomas Lukaschewski, Muriel Reuter, Annika Rips, and Carina Zöllner for help with data acquisition. We thank Lorena Deuker for support with data analysis and Sebastian Ocklenburg for thoughtful advice on the lateralization of our amygdala effects. N.A. was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—project nos. 316803389 – SFB 1280 "Extinction learning," 122679504 – SFB 874 "Integration and representation of sensory processes," and 419049386 – FOR 2812 "Construction scenarios of the past." O.T.W. was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—project nos. 122679504 – SFB 874 "Integration and representation of sensory processes" and 419039274 – FOR 2812 "Construction scenarios of the past."

## AUTHOR CONTRIBUTIONS

A.B., N.A., and O.T.W. developed key hypotheses. A.B., M.-C.F., O.T.W., and N.A. designed the study. A.B. collected and analyzed data with support from M.-C.F., R.H., O.T.W., and N.A. A.B. and N.A. wrote the manuscript, with substantial support from all authors.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 6, 2021
Revised: August 27, 2021
Accepted: September 16, 2021
Published: October 14, 2021

## REFERENCES

1. Wolf, O.T. (2019). Memories of and influenced by the Trier Social Stress Test. Psychoneuroendocrinology *105*, 98–104.

Please cite this article in press as: Bierbrauer et al., The memory trace of a stressful episode, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.09.044



2. Joëls, M., Pu, Z., Wiegert, O., Oitzl, M.S., and Krugers, H.J. (2006). Learning under stress: how does it work? Trends Cogn. Sci. *10*, 152–158.
3. Joëls, M., Fernandez, G., and Roozendaal, B. (2011). Stress and emotional memory: a matter of timing. Trends Cogn. Sci. *15*, 280–288.
4. Roozendaal, B., McEwen, B.S., and Chattarji, S. (2009). Stress, memory and the amygdala. Nat. Rev. Neurosci. *10*, 423–433.
5. Shields, G.S., Sazma, M.A., McCullough, A.M., and Yonelinas, A.P. (2017). The effects of acute stress on episodic memory: a meta-analysis and integrative review. Psychol. Bull. *143*, 636–675.
6. Wiemers, U.S., Sauvage, M.M., Schoofs, D., Hamacher-Dang, T.C., and Wolf, O.T. (2013). What we remember from a stressful episode. Psychoneuroendocrinology *38*, 2268–2277.
7. Mather, M. (2007). Emotional arousal and memory binding: an object-based framework. Perspect. Psychol. Sci. *2*, 33–52.
8. Kriegeskorte, N., and Kievit, R.A. (2013). Representational geometry: integrating cognition, computation, and the brain. Trends Cogn. Sci. *17*, 401–412.
9. Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. *2*, 4.
10. LaRocque, K.F., Smith, M.E., Carr, V.A., Witthoft, N., Grill-Spector, K., and Wagner, A.D. (2013). Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. J. Neurosci. *33*, 5466–5474.
11. Brunec, I.K., Robin, J., Olsen, R.K., Moscovitch, M., and Barense, M.D. (2020). Integration and differentiation of hippocampal memory traces. Neurosci. Biobehav. Rev. *118*, 196–208.
12. Visser, R.M., Scholte, H.S., Beemsterboer, T., and Kindt, M. (2013). Neural pattern similarity predicts long-term fear memory. Nat. Neurosci. *16*, 388–390.
13. Kirschbaum, C., Pirke, K.-M., and Hellhammer, D.H. (1993). The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology *28*, 76–81.
14. Wiemers, U.S., Schoofs, D., and Wolf, O.T. (2013). A friendly version of the trier social stress test does not activate the HPA axis in healthy men and women. Stress *16*, 254–260.
15. Herten, N., Otto, T., and Wolf, O.T. (2017). The role of eye fixation in memory enhancement under stress - an eye tracking study. Neurobiol. Learn. Mem. *140*, 134–144.
16. Bakker, A., Kirwan, C.B., Miller, M., and Stark, C.E.L. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. Science *319*, 1640–1642.
17. Yassa, M.A., and Stark, C.E.L. (2011). Pattern separation in the hippocampus. Trends Neurosci. *34*, 515–525.
18. Axmacher, N., Elger, C.E., and Fell, J. (2008). Memory formation by refinement of neural representations: the inhibition hypothesis. Behav. Brain Res. *189*, 1–8.
19. Xue, G. (2018). The neural representations underlying human episodic memory. Trends Cogn. Sci. *22*, 544–561.
20. Yonelinas, A.P., and Ritchey, M. (2015). The slow forgetting of emotional episodic memories: an emotional binding account. Trends Cogn. Sci. *19*, 259–267.
21. Fusi, S., Miller, E.K., and Rigotti, M. (2016). Why neurons mix: high dimensionality for higher cognition. Curr. Opin. Neurobiol. *37*, 66–74.
22. McKenzie, S., Frank, A.J., Kinsky, N.R., Porter, B., Rivière, P.D., and Eichenbaum, H. (2014). Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. Neuron *83*, 202–215.
23. Wood, E.R., Dudchenko, P.A., Robitsek, R.J., and Eichenbaum, H. (2000). Hippocampal neurons encode information about different types of memory episodes occurring in the same location. Neuron *27*, 623–633.
24. Teyler, T.J., and Rudy, J.W. (2007). The hippocampal indexing theory and episodic memory: updating the index. Hippocampus *17*, 1158–1169.
25. Liu, X., Ramirez, S., Pang, P.T., Puryear, C.B., Govindarajan, A., Deisseroth, K., and Tonegawa, S. (2012). Optogenetic stimulation of a hippocampal engram activates fear memory recall. Nature *484*, 381–385.
26. Han, J.H., Kushner, S.A., Yiu, A.P., Hsiang, H.L., Buch, T., Waisman, A., Bontempi, B., Neve, R.L., Frankland, P.W., and Josselyn, S.A. (2009). Selective erasure of a fear memory. Science *323*, 1492–1496.
27. Saez, A., Rigotti, M., Ostojic, S., Fusi, S., and Salzman, C.D. (2015). Abstract context representations in primate amygdala and prefrontal cortex. Neuron *87*, 869–881.
28. Gothard, K.M. (2020). Multidimensional processing in the amygdala. Nat. Rev. Neurosci. *21*, 565–575.
29. Rutishauser, U., Tudusciuc, O., Neumann, D., Mamelak, A.N., Heller, A.C., Ross, I.B., Philpott, L., Sutherling, W.W., and Adolphs, R. (2011). Single-unit responses selective for whole faces in the human amygdala. Curr. Biol. *21*, 1654–1660.
30. Wang, S., Tudusciuc, O., Mamelak, A.N., Ross, I.B., Adolphs, R., and Rutishauser, U. (2014). Neurons in the human amygdala selective for perceived emotion. Proc. Natl. Acad. Sci. USA *111*, E3110–E3119.
31. Graner, J.L., Stjepanović, D., and LaBar, K.S. (2020). Extinction learning alters the neural representation of conditioned fear. Cogn. Affect. Behav. Neurosci. *20*, 983–997.
32. Jin, J., Zelano, C., Gottfried, J.A., and Mohanty, A. (2015). Human amygdala represents the complete spectrum of subjective valence. J. Neurosci. *35*, 15145–15156.
33. Murty, V.P., Ritchey, M., Adcock, R.A., and LaBar, K.S. (2010). fMRI studies of successful emotional memory encoding: a quantitative meta-analysis. Neuropsychologia *48*, 3459–3469.
34. Roozendaal, B., and McGaugh, J.L. (2011). Memory modulation. Behav. Neurosci. *125*, 797–824.
35. Ocklenburg, S., Korte, S.M., Peterburs, J., Wolf, O.T., and Güntürkün, O. (2016). Stress and laterality - the comparative perspective. Physiol. Behav. *164* (Pt A), 321–329.
36. Baas, D., Aleman, A., and Kahn, R.S. (2004). Lateralization of amygdala activation: a systematic review of functional neuroimaging studies. Brain Res. Brain Res. Rev. *45*, 96–103.
37. Dickerson, S.S., and Kemeny, M.E. (2004). Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychol. Bull. *130*, 355–391.
38. Miller, R., Plessow, F., Kirschbaum, C., and Stalder, T. (2013). Classification criteria for distinguishing cortisol responders from nonresponders to psychosocial stress: evaluation of salivary cortisol pulse detection in panel designs. Psychosom. Med. *75*, 832–840.
39. Tulving, E. (1985). Memory and consciousness. Can. Psychol. *26*, 1–12.
40. Cohen, N.J., and Eichenbaum, H. (1993). Memory, Amnesia, and the Hippocampal System (MIT Press).
41. Sazma, M.A., McCullough, A.M., Shields, G.S., and Yonelinas, A.P. (2019). Using acute stress to improve episodic memory: the critical role of contextual binding. Neurobiol. Learn. Mem. *158*, 1–8.
42. Burgess, N., Maguire, E.A., and O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. Neuron *35*, 625–641.
43. Campbell, J., and Ehlert, U. (2012). Acute psychosocial stress: does the emotional stress response correspond with physiological responses? Psychoneuroendocrinology *37*, 1111–1134.
44. Inman, C.S., Manns, J.R., Bijanki, K.R., Bass, D.I., Hamann, S., Drane, D.L., Fasano, R.E., Kovach, C.K., Gross, R.E., and Willie, J.T. (2018). Direct electrical stimulation of the amygdala enhances declarative memory in humans. Proc. Natl. Acad. Sci. USA *115*, 98–103.
45. Cahill, L., Babinsky, R., Markowitsch, H.J., and McGaugh, J.L. (1995). The amygdala and emotional memory. Nature *377*, 295–296.
46. Qin, S., Hermans, E.J., van Marle, H.J.F., and Fernández, G. (2012). Understanding low reliability of memories for neutral information encoded under stress: alterations in memory-related activation in the hippocampus and midbrain. J. Neurosci. *32*, 4032–4041.

Please cite this article in press as: Bierbrauer et al., The memory trace of a stressful episode, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.09.044



47. Bahtiyar, S., Gulmez Karaca, K., Henckens, M.J.A.G., and Roozendaal, B. (2020). Norepinephrine and glucocorticoid effects on the brain mechanisms underlying memory accuracy and generalization. Mol. Cell. Neurosci. *108*, 103537.

48. Sergerie, K., Chochol, C., and Armony, J.L. (2008). The role of the amygdala in emotional processing: a quantitative meta-analysis of functional neuroimaging studies. Neurosci. Biobehav. Rev. *32*, 811–830.

49. Wright, C.I., Fischer, H., Whalen, P.J., McInerney, S.C., Shin, L.M., and Rauch, S.L. (2001). Differential prefrontal cortex and amygdala habituation to repeatedly presented emotional stimuli. Neuroreport *12*, 379–383.

50. Gläscher, J., and Adolphs, R. (2003). Processing of the arousal of subliminal and supraliminal emotional stimuli by the human amygdala. J. Neurosci. *23*, 10274–10282.

51. Phillips, M.L., Medford, N., Young, A.W., Williams, L., Williams, S.C.R., Bullmore, E.T., Gray, J.A., and Brammer, M.J. (2001). Time courses of left and right amygdalar responses to fearful facial expressions. Hum. Brain Mapp. *12*, 193–202.

52. Rohleder, N., and Nater, U.M. (2009). Determinants of salivary alpha-amylase in humans and methodological considerations. Psychoneuroendocrinology *34*, 469–485.

53. Lorentz, K., Gütschow, B., and Renner, F. (1999). Evaluation of a direct α-amylase assay using 2-chloro-4-nitrophenyl-α-D-maltotrioside. Clin. Chem. Lab. Med. *37*, 1053–1062.

54. Heatherton, T.F., and Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. J. Pers. Soc. Psychol. *60*, 895–910.

55. Laux, L., Glanzmann, P., Schaffner, P., and Spielberger, C.D. (1981). Das State-Trait-Angstinventar. Theoretische Grundlagen und Handanweisung (Beltz Test).

56. Merten, J., and Krause, R. (1993). Differentielle Affekt-Skala: DAS (Univ., Fachbereich Sozial- und Umweltwiss., Fachrichtung Psychologie).

57. Watson, D., Clark, L.A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. J. Pers. Soc. Psychol. *54*, 1063–1070.

58. Hautzinger, M., Keller, F., and Kühner, C. (2006). BDI-II Beck Depression-Inventar, Second Edition (Harcourt Test Services).

59. Deuker, L., Bellmund, J.L., Navarro Schröder, T., and Doeller, C.F. (2016). An event map of memory space in the hippocampus. eLife *5*, 1–26.

60. Theves, S., Fernandez, G., and Doeller, C.F. (2019). The hippocampus encodes distances in multidimensional feature space. Curr. Biol. *29*, 1226–1231.e3.

61. Ehlers, A., and Clark, D.M. (2000). A cognitive model of posttraumatic stress disorder. Behav. Res. Ther. *38*, 319–345.

62. R Core Team (2018). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

63. Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. J. Stat. Softw. *67*, 1–48.

64. Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., and Jensen, S.P. (2020). lmerTest.

65. Lenth, R. (2019). emmeans: estimated marginal means, aka least-squares means.

66. Snodgrass, J.G., and Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. J. Exp. Psychol. Gen. *117*, 34–50.

67. Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., et al. (2004). Automatically parcellating the human cerebral cortex. Cereb. Cortex *14*, 11–22.

68. Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage *31*, 968–980.

69. Shen, J. (2021). Tools for NIfTI and ANALYZE image.

70. Dimsdale-Zucker, H.R., and Ranganath, C. (2018). Representational similarity analyses: a practical guide for functional MRI applications. In Handbook of Behavioral Neuroscience (Elsevier), pp. 509–525.

71. Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. J. Exp. Soc. Psychol. *49*, 764–766.

72. Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences, Second Edition (Lawrence Erlbaum Associates).

73. Selya, A.S., Rose, J.S., Dierker, L.C., Hedeker, D., and Mermelstein, R.J. (2012). A practical guide to calculating Cohen's f(2), a measure of local effect size, from PROC MIXED. Front. Psychol. *3*, 111.

74. Lorah, J. (2018). Effect size measures for multilevel models: definition, interpretation, and TIMSS example. Large-Scale Assess. Educ. *6*, 8.

75. Ben-Shachar, M.S., Ludecke, D., and Makowski, D. (2021). effectsize: estimation of effect size indices and standardized parameters. J. Open Source Soft. *5*, 2815.

Please cite this article in press as: Bierbrauer et al., The memory trace of a stressful episode, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.09.044



CellPress
OPEN ACCESS

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Deposited data | | |
| Scripts | https://osf.io/B97PD | https://doi.org/10.17605/OSF.IO/B97PD |
| Preprocessed data | https://osf.io/B97PD | https://doi.org/10.17605/OSF.IO/B97PD |
| Software and algorithms | | |
| MATLAB 2018a | https://www.mathworks.com | RRID: SCR_001622 |
| Presentation v18.0 | https://www.neurobs.com | RRID: SCR_002521 |
| Unreal Engine 4.11 | https://www.unrealengine.com/en-US | N/A |
| Freesurfer v6.0.0 | https://surfer.nmr.mgh.harvard.edu | RRID: SCR_001847 |
| SPM 12 | http://www.fil.ion.ucl.ac.uk/spm | RRID: SCR_007037 |
| R 3.5.0 | http://www.r-project.org | RRID: SCR_001905 |

### RESOURCE AVAILABILITY

#### Lead Contact

Further information and request for resources and should be directed to and will be fulfilled by the lead contact, Anne Bierbrauer (anne.bierbrauer@rub.de).

#### Materials Availability

This study did not generate new materials.

#### Data and code availability

All data and original code has been deposited at Open Science Framework and is publicly available as of the date of publication. DOIs are listed in the key resources table. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

#### Participants

We tested 65 healthy, adult, right-handed males at 19-33 years of age (for allocation to experimental groups, see Figure 1A; Table S1). We excluded participants with neurological or psychiatric diseases, medical treatment that could potentially influence the HPA axis, previous participation in the Trier Social Stress Test (TSST) or one of the pilot studies, a body mass index < 18 or > 30, smoking, no fluency in German language, and MR incompatibility (due to metal implants, tinnitus, or claustrophobia). All participants were students of Ruhr University Bochum, gave their informed consent, and received monetary compensation of 50€ or study credits. The local ethics committee approved the study (No. 15/191), and the study was conducted in accordance with the declaration of Helsinki. One participant was excluded post hoc due to an incidental finding in the MR scan. The sample thus included 64 participants [age (mean ± SD): 24.05 ± 3.27].

### METHOD DETAILS

#### Hormonal assessment

We instructed participants to refrain from excessive sports and alcohol consumption or medication the day before testing. Moreover, participants should not eat or drink anything but water 1 h before testing. We used Salivettes (Sarstedt, Germany) to acquire saliva samples at four time points on the first testing day and once at the beginning of the second testing day. Cortisol was analyzed by immunoassay (IBL, Hamburg, Germany) and alpha-amylase (sAA) by a quantitative enzyme kinetic method as described in detail elsewhere.[52,53]

Please cite this article in press as: Bierbrauer et al., The memory trace of a stressful episode, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.09.044



## Questionnaires

In a custom sleep and medication questionnaire, we asked participants for get-up time, h of sleep, if they had had breakfast, breakfast time, as well as medication intake on the day before. Participants rated their current self-esteem on the "State Self-Esteem Scale" (SSES; five-point scale; 20 items).[54] Furthermore, they rated their trait anxiety and rated their state anxiety before and after the stress procedure on the German version of the "State-Trait Anxiety Inventory" (STAI; four-point scale; 40 items).[55] On the German version of the "Differential Emotion Scale" ("Differentielle Affekt-Skala," DAS),[56] they rated their current affect on different dimensions before and after the stress procedure. Using the "Positive and Negative Affect Scale" (PANAS; five-point scale; 20 items),[57] we assessed the current affect of the participants before and after the stress procedure. Finally, we used Beck's Depression Inventory II (BDI II; 21 items)[58] to check for symptoms of depression. Only PANAS and SSES scores were analyzed before and after the (f-)TSST in order to assess if the TSST indeed induced subjective stress (operationalized by higher negative affect and lower state self-esteem). All other questionnaires but the DAS were included to rule out significant *a priori* differences between the experimental groups (Table S1). The DAS was included to assess the emotions induced by the (f-)TSST in more detail, but it is not part of the current study.

## Stress procedure and stress response

### *Stress procedure*

The Trier Social Stress Test (TSST) is an ecologically valid and reliable procedure to induce psychosocial stress.[13] We used a customized version that has been described in detail in previous studies.[6,14] Participants first filled in a sham questionnaire to increase ego-threat and prepared their free speech (5 min). Afterward, we instructed participants to speak freely in a sham job interview in front of a committee for 10 min. The committee consisted of a male and a female member acting in a neutral and reserved manner. Participants were video- and audiotaped during the speech.

Wiemers et al. developed a control condition, the f-TSST, that matched central characteristics of the TSST without inducing stress.[14] In the f-TSST, participants filled in a questionnaire about their school years, university track, career aspirations, hobbies, and favorite book or movie (5 min). Afterward, participants were instructed to hold a free speech about their life and career aspirations (10 min). The committee consisted of a male and a female member acting friendly by nodding and smiling. There was no video- and audiotaping.

### *Cortisol response*

Saliva cortisol levels were analyzed using a mixed model ANOVA with "time" [baseline, i.e., before (f-)TSST, versus 1min, 10min, and 25min after (f)TSST] as within-subject factor and "experimental condition" (TSST versus f-TSST) as between-subject factor (Figure 2A). *P*-values of the follow-up simple effects tests were Bonferroni corrected for the number of comparisons (i.e., 4 time points or 2 experimental conditions).

Mean cortisol concentration was log-transformed for statistical analysis, as cortisol values were not normally distributed. Cortisol levels changed as a function of time ($F(3,186) = 15.75$, $p < 0.001$) and, more importantly, depended on experimental condition, as indicated by a significant interaction ($F(3,186) = 9.10$, $p < 0.001$). Compared to baseline, cortisol was significantly elevated 1 min and 10 min after the experimental intervention only in the TSST condition (both $t(186) \geq 3.49$, both $P_{Bonferroni} \leq 0.004$), but not in the f-TSST condition (both $t(186) \leq 0.22$, both $P_{Bonferroni} \approx 1$). 25 min after the experimental intervention, cortisol in the TSST condition did not differ from baseline ($t(186) = 2.07$, $P_{Bonferroni} = 0.242$) and decreased significantly in the f-TSST condition ($t(186) = 3.88$, $P_{Bonferroni} < 0.001$), most likely reflecting the typical circadian decrease over the course of the day. Cortisol concentrations were significantly higher in TSST as compared to f-TSST 10 min and 25 min after the experimental intervention (both $t(79.6) \geq 2.94$, both $P_{Bonferroni} \leq 0.017$), but not 1 min afterward ($t(79.6) = 1.93$, $P_{Bonferroni} = 0.230$). Thus, participants experienced a significant increase in cortisol only after TSST, but not after f-TSST.

### *Subjective stress response*

We analyzed three measures of subjective emotional response, i.e., the mean scores of negative and positive affect of the PANAS scale, and the state self-esteem mean score of the SSES scale. Data was analyzed with a mixed model ANOVA with "time" [before versus after (f-)TSST] as within-subject factor and "experimental condition" (TSST versus f-TSST) as between-subject factor (Figure 2B). *P*-values of the follow-up simple effects tests were Bonferroni corrected for the number of comparisons (i.e., 2 times or 2 experimental conditions).

There was no main effect of time for the negative affect and for the state self-esteem scores (both $F(1,62) \leq 0.52$, both $P \geq 0.474$). Positive affect increased significantly after the experimental manipulation ($F(1,62) = 34.12$, $p < 0.001$). We found a significant interaction in all subjective emotional response measures (all $F(1,62) \geq 4.14$, all $P \leq 0.046$), indicating a subjective emotional response depending on the experimental condition. In the stressful condition, negative affect increased ($t(62) = -2.77$, $P_{Bonferroni} = 0.015$), whereas there was no difference for participants in the non-stressful condition ($t(62) = 1.73$, $P_{Bonferroni} = 0.178$). Positive affect increased in both conditions, but, as indicated by the significant interaction, it did even more so in the non-stressful condition (TSST: $t(62) = -2.73$, $P_{Bonferroni} = 0.016$, f-TSST: $t(62) = -5.48$, $P_{Bonferroni} < 0.001$). Finally, state self-esteem was not affected in the TSST condition ($t(62) = 1.67$, $P_{Bonferroni} = 0.200$), but it increased significantly in the f-TSST condition ($t(62) = -2.63$, $P_{Bonferroni} = 0.022$). Taken together, we found evidence for higher negative affect specifically after the TSST and stronger increases of positive affect and self-esteem after the f-TSST.

## Memory items: Objects and faces

We arranged 24 items in the experimental intervention room, 20 items on the table in front of the committee (e.g., a stapler or a puncher, see Figures 1A and S1B), and four items in the room behind and next to the committee (e.g., a trash bin). During the

Please cite this article in press as: Bierbrauer et al., The memory trace of a stressful episode, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.09.044



CellPress
OPEN ACCESS

interview, the committee members interacted with a subset of the objects in accordance with a script. For instance, the male committee member took the cola can, filled the cup, and drank a sip at minute one during the interview (for a depiction of all actions, see Figure S1A).

### Counterbalancing

We counterbalanced central objects across participants, i.e., half of the participants in each experimental condition watched the committee manipulating a different subset of objects (version A: n = 16 in TSST, n = 16 in f-TSST; version B: n = 17 in TSST, n = 15 in f-TSST; test for equal distribution of participants: $\chi^2(1) = 0.06$, $P \approx 1$). In version A, the committee manipulated the tea, the mug, the teapot, the paper tissue, the trash bin, the pencil sharpener, the hand lotion, the stapler, and the filing. In version B, the committee manipulated the cola can, the plastic cup, the candy, the rubber, the highlighter, the sticky tape, the puncher, and the folder. Clipboards, pencils, and the stopwatch were always central, while TV, ventilator, and book were never central. We needed these non-counterbalanced, central objects for the (f-)TSST procedure (e.g., the stopwatch was always needed to measure the time during the experiment) and we added the same amount of non-counterbalanced, peripheral objects. We excluded the non-counterbalanced objects from all analyses. For a description of the control experiment to ensure that there were no pre-existing differences between counterbalanced object groups, see below.

For the recognition memory test and the fMRI experiment, we added 48 distractor objects, 24 of which were lures for each of the objects in the room, differing from the previously presented objects only in shape or color (e.g., a red stapler as a lure for the blue stapler). The other 24 distractors were completely different objects that had not been in the room. We chose objects that could have potentially been in the room, like a clock or a cookie box. Moreover, we added the faces of the committee members together with 2 distractor faces (one male and one female face). This resulted in 76 pictures presented during the recognition memory task and the fMRI experiment.

### Control experiment

In the current study, we characterized memory representations one day after the stressful episode, which allows for conclusions about the structure of the memory trace after it has been established. Several previous studies on representational changes employed pre-post designs to investigate the effect of interventions on representational similarities (e.g., Deuker et al.[59] and Theves et al.[60]). This allowed them to control for the pre-existing similarity structure between items, e.g., due to visual or semantic features. We chose not to measure pre-existing similarities for several reasons: First, showing the objects before the (f-)TSST may have obscured stress effects on memory, since attention of the participants would have been drawn to these objects. Second, the subsequent memory test would have become confusing and difficult to interpret, because participants would have needed to indicate whether they remembered the objects from the actual (f-)TSST or from the preceding scanning session.

We excluded in a different way that our results were affected by possible pre-existing differences in the similarity structures of central versus peripheral objects: We counterbalanced central and peripheral objects: One group of participants conducted the study with episode A in which one half of the objects was central and the other half peripheral; another group underwent episode B with the other half of the objects being central. This procedure ensured that differential stress effects on central versus peripheral objects are not due to pre-existing differences between the object types.

In addition, we matched the objects of episodes A and B with respect to their neural representations based on a control experiment: N = 14 participants underwent the fMRI task without completing the (f-)TSST on the previous day. By analyzing neural representations of the objects in this control group, we obtained a measure of *a priori* relatedness of the objects, presumably mostly based on visual and semantic similarity. Object group A did not differ from object group B in terms of the representational measures. Representational reliability (all $t(13) \leq 1.39$, all $P_{Bonferroni} \geq 0.755$), similarity difference score (all $t(13) \leq 2.14$, all $P_{Bonferroni} \geq 0.208$), object similarity (all $t(13) \leq 1.63$, all $P_{Bonferroni} \geq 0.512$), and object-distractor similarity (all $t(13) \leq 0.51$, all $P_{Bonferroni} \approx 1$) did not differ in any of the four ROIs. We did not show the stressor pictures in the control experiment, as there was no stress intervention on the previous day.

Furthermore, a control analysis in the main sample of participants confirmed that there were no differences between the object groups A and B. Object group A did not differ from object group B in terms of representational reliability (all $t(63) \leq 1.80$, all $P_{Bonferroni} \geq 0.309$), similarity difference score (all $t(63) \leq 1.22$, all $P_{Bonferroni} \geq 0.906$), object similarity (all $t(63) \leq 1.28$, all $P_{Bonferroni} \geq 0.825$), object-distractor similarity (all $t(63) \leq 1.52$, all $P_{Bonferroni} \geq 0.531$), stressor-object similarity (all $t(63) \leq 2.10$, all $P_{Bonferroni} \geq 0.160$) in any of the four ROIs. Taken together, object groups did not differ in terms of representational measures in the control experiment and in the main experiment.

## Memory assessment

We assessed memory for the objects and the faces of the committee members that occurred in the (f-)TSST. Participants performed three memory tests, i.e., a free recall test, a recognition memory test, and a spatial memory test.

### Free recall

We instructed participants to write down everything (i.e., all objects) they could remember from the experimental intervention room. There was no time limit and participants were instructed to let the experimenter know when they could not remember any more details. Participants took ~5 min to complete this task.

### Recognition memory

We presented the recognition memory task using Presentation software (Version 18.0, Neurobehavioral Systems, Berkeley, CA, https://www.neurobs.com/) on a laptop with a resolution of 1920 × 1080 pixels and a screen diagonal of 35.6 cm. Participants viewed the 76 pictures of objects and faces and rated for each picture their confidence of having seen it on the previous day. We used a six

Please cite this article in press as: Bierbrauer et al., The memory trace of a stressful episode, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.09.044



point rating scale ranging from 1 (very sure of having seen the object on the previous day) to 6 (very sure of not having seen the object on the previous day). Trials were presented with a 1 s inter-trial-interval and no time limit for the response (i.e., the next trial started after the response). Participants took 7 min 54.38 s ± 1 min 7.37 s (mean ± SD) to complete this task and stressed participants did not differ from controls ($t(62) = -0.87$, $p = 0.389$).

#### *Spatial memory*

We implemented the spatial memory task using Unreal Engine (Epic Games, version 4.11) on a laptop with a screen diagonal of 45 cm, a resolution of 1920 × 1080 pixels and a frame rate of 60 frames/s. Participants saw a virtual version of the table that was in front of the committee members during the experimental intervention. The table was empty and the objects that had been on the table were placed left and right of the table. We instructed participants to drag and drop the objects to the correct locations on the table. Participants could rearrange the objects until the table corresponded to their memory. We instructed them to guess the location of objects they could not remember until all objects were on the table. The task was self-paced and participants took 2 min 36.99 s ± 1 min 3.80 s (mean ± SD) to complete this task. Stressed participants were significantly faster than controls [$t(61) = -2.15$, $p = 0.035$; TSST: 2 min 20.36 s ± 9.11 s (mean ± SEM); f-TSST: 2 min 54.28 s ± 12.95 s (mean ± SEM)]. On participant could not complete this task due to technical issues.

### FMRI data acquisition

We conducted the fMRI recordings at the Bergmannsheil hospital in Bochum using a 3 T Philips Achieva scanner (Best, the Netherlands) with a 32-channel headcoil. First, we acquired a high-resolution whole-brain structural brain scan using a T1-weighted sequence at 1 mm isotropic resolution (FOV: 240 mm x 240 mm, 220 transversally oriented slices). The total acquisition time (TA) was 6 min 2 s. Second, blood oxygenation level-dependent (BOLD) contrast images were acquired with a T2*-weighted gradient echo EPI sequence with 2.5-mm isotropic resolution (TR = 2500 ms, TE = 30 ms, FA = 90°, FOV = 96 mm x 96 mm, 46 transversal slices in ascending order without slice gap). The TA for each of the functional runs was 16 min 59.22 s ± 54.96 s (mean ± SD; corresponding to 408 ± 22 scans; mean ± SD). We discarded the first five images of each session to allow for signal steady-state transition.

Participants viewed the experiment via MR-compatible liquid crystal display (LCD) goggles (VisuaStim Digital, Resonance Technology, Northridge, CA, USA) with a resolution of 800 × 600 pixels. They performed the task using an MR-compatible keyboard with the right index finger.

We presented the experiment using Presentation software (Version 18.0, Neurobehavioral Systems, Berkeley, CA, https://www.neurobs.com/). Participants viewed the same 76 pictures of objects as in the recognition memory task. Every item was presented for 1500 ms, followed by an inter-trial interval of 3750 ms or 5750 ms (uniformly randomly distributed). In each run, every item was shown twice: In the first half of the run, all 76 items were shown in random order, followed by a pause of 30 s in the middle of the run. In the second half of the run, all 76 items were shown again in random order. We used an oddball cover task to make sure that participants paid attention to the items during the fMRI scan: On 5% of the pictures (i.e., 8 pictures per run), we presented a small fly on the object for 400 ms. We instructed participants to press the response button as fast as possible when they encountered an item with a fly. At the end of each run, we provided feedback to the participants about the amount of flies identified correctly.

Even though we were interested in memory representations we did not employ an explicit retrieval task inside the scanner for several reasons: First, an explicit task would have interfered with our aim to present every object 12 times. Second, one may argue that using an implicit retrieval task increases the ecological validity of our paradigm, since stressful or traumatic experiences are rarely retrieved explicitly – instead, they are typically triggered by external cues that involuntarily bring back their memory.[61] Previous studies have successfully employed a passive viewing task to assess memory representations and their transformation (e.g., Deuker et al.[59]).

Participants took 16 min 59.22 s ± 54.96 s (mean ± SD) to complete one of the functional runs. Sixty-two participants completed 6 functional runs, 3 participants completed 5 functional runs.

In a control experiment, we acquired data from 18 healthy, adult, right-handed participants of which we had to exclude four participants due to technical issues. The control sample thus comprised 14 participants [12 males, age (mean ± SD): 24.90 ± 2.69] who completed the same fMRI task as the main sample without participating in an (f-)TSST. As in the main experiment, fMRI data was acquired in two sessions, but we scheduled the sessions on two consecutive days instead of a single day. We excluded participants with neurological or psychiatric diseases, medical treatment that could potentially influence the HPA axis, a body mass index < 18 or > 30, and MR incompatibility (due to metal implants, tinnitus, or claustrophobia). All control participants gave their informed consent, and received monetary compensation of 30€ or study credits. The local ethics committee approved the study (No. 15/191), and the study was conducted in accordance with the declaration of Helsinki.

### Experimental procedure

The experiment was performed on two consecutive days (Figure 1A). On the first day, the participant met the experimenter in the preparation room at Ruhr University Bochum either at 9:30 or at 11:00. The experimenter explained the procedure of both days, and the participant gave his informed consent. The experimenter provided participants of both experimental conditions with the same information about the experimental intervention, i.e., that they would have to do an interview in front of two other persons, being filmed and audiotaped. Participants did not receive any further information about the details of the situation or the content of the interview. Afterward, participants filled in the questionnaires "Sleep," SSES, STAI ("state" and "trait"), DAS, and PANAS and the experimenter took a saliva sample (BL). Then, the experimenter took the participant to the room in which the experimental intervention took place. After entering this room, the participant received detailed information about the procedure, corresponding either to the TSST

Please cite this article in press as: Bierbrauer et al., The memory trace of a stressful episode, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.09.044



CellPress
OPEN ACCESS

or to the f-TSST condition. The experimenter left the room and came back 15 min later to pick up the participant. Immediately after the experimental intervention, the participant provided a second saliva sample (+1). Afterward, experimenter and participant went back to the preparation room and the participant filled in the questionnaires PANAS, SSES, and STAI ("state"). 10 min after the end of the experimental intervention, a third saliva sample was taken (+10). Finally, the participant filled in the questionnaires DAS and BDI-II and gave a fourth saliva sample (+25), which constituted the end of the first day. The testing procedure on day 1 took 75 min.

On the second day, the participant met the experimenter at the MR room of Bergmannsheil hospital in Bochum in the afternoon. First, he filled in the PANAS questionnaire and provided a final saliva sample (fMRI). Afterward, he performed the free recall and the recognition memory test. Then, the participant was taken to the scanner, a T1 scan was acquired (6 min), and afterward, the participant performed three out of six experimental runs, each taking ~16 min. After ~50 min, the experimenter took the participant out of the scanner for a short break of about 10-15 min. After the break, the participant went back into the scanner for three more experimental runs (again ~16 min for each run). After the end of the fMRI experiment, participants performed the spatial memory task. Spatial memory tests of the first seven participants were taken before the fMRI experiment. The testing procedure on day 2 took 150 min.

## QUANTIFICATION AND STATISTICAL ANALYSIS

We extracted behavioral data from the logfiles using MATLAB (2018a, The MathWorks, Massachusetts). Regions of interest (ROIs) were created using FreeSurfer (v6.0.0). We used SPM12 (https://www.fil.ion.ucl.ac.uk/spm/) for all fMRI analyses and statistics were done in R[62] (3.5.0) using the lme4[63] (v1.1-17), lmerTest[64] (v3.1-3), and the emmeans[65] (v1.2.2) packages.

### Behavioral analyses

We calculated mean scores of get-up time and h of sleep (Sleep and medication questionnaire), of positive (PA) and negative (NA) affect scores (PANAS), of the trait anxiety and of the state anxiety scores (STAI), as well as of the SSES score before the experimental intervention (and on day 2 for PA and NA). We compared the mean ratings between the experimental groups to ensure that there were no *a priori* differences (Table S1).

Further, we calculated PA, NA and SSES mean scores after the stress procedure to analyze the differences in subjective stress response between experimental conditions. Please note that, even though the SSES manual suggests calculating sum scores, we used a mean score to ensure that the different subjective rating scales have a similar range.

Free recall performance was assessed in terms of the number of "experimental" objects that participants produced in each object category (i.e., central and peripheral). "Experimental" here refers to the objects that were also part of the recognition memory and of the fMRI experiment (see section "Memory items: Objects and faces"). For example, "teapot" was an experimental object; "table" did not count as an experimental object.

We used the same procedure as Wiemers et al. (2013) to analyze object recognition performance.[6] We dichotomized the participants's answers into "yes, seen the object during the experimental intervention" and "no, not seen the object during the experimental intervention." We then calculated "Pr" (hit rate – false alarm rate), a memory performance measure according to the Two-High Threshold Model[66] for each object category (i.e., central and peripheral).

In order to characterize the recognition memory comprehensively, we plotted receiver operating characteristic (ROC) curves to both easy and difficult distractors (Figure S2D; Table S3). This allows for an in-depth understanding of the response behavior of the participants and provides a more fine-grained picture of the false alarms that were provoked by difficult or easy distractors. We calculated the area under the curve (AUC) as a measure of memory performance.

Generally, we only used counterbalanced objects (and corresponding difficult distractors) for calculations of hit and false alarm rates. For calculation of false alarm rates, we considered false alarms to easy distractors (that were not matched one-by-one to individual objects) and to difficult distractors that matched the individual objects. For example, the false alarm rate for central objects was calculated using all easy distractors plus those difficult distractors that corresponded to the central objects.

Overall hit and false alarm rates are an average across all objects from a given experimental condition, while we were interested in the representational similarity between individual objects in the amygdala. In other words, we were interested in the representational structures that made one object more memorable than another one. Modeling the data on the level of individual objects allowed us to establish a link between the object-specific representational measure (e.g., stressor-object similarity) and the memory strength for this object. Object-specific memory strength was defined as the rating for each object minus the participant-wise mean rating of all objects, which corrected for participant-wise response biases. We inverted the values (simply by subtracting them from 7), such that higher values reflected higher memory strength.

We assessed spatial memory performance as the "distance error," which refers to the Euclidean distance between the correct and the presumed object location as indicated by the participant. Mean distance error was calculated separately for each object category (i.e., central and peripheral).

### Extracting ROIs with Freesurfer

We created anatomical ROIs from structural scans using the automatic parcellation procedure as implemented in FreeSurfer v6.0 (https://surfer.nmr.mgh.harvard.edu/). In short, the procedure included intensity normalization, registration to Talairach space, skull stripping, white matter segmentation, tessellation of the white matter boundaries, and automatic correction of topological defects.[67] Segmented brain images were parcellated into cortical and subcortical regions according to the Desikan-Killiany atlas.[68] From this

Please cite this article in press as: Bierbrauer et al., The memory trace of a stressful episode, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.09.044



parcellation, we derived the ROIs for left and right amygdala and hippocampus (no overlapping voxels between the ROIs). ROI masks were coregistered to functional space to extract voxels from the functional scans in native space.

## FMRI analyses

We excluded functional runs due to head motion of more than the voxel size (n = 29 runs), if participants did not pay attention and missed more than two flies in the oddball task (n = 13 runs), or because of technical issues (n = 5 runs). After exclusion, there were 5.30 ± 1.05 (mean ± SD) functional runs for each participant. The exclusion of an fMRI run based on the criteria described above did never result in the loss of an entire dataset. We thus analyzed 64 participants: n = 1 with two runs, n = 6 with three runs, n = 4 with four runs, n = 15 with five runs, and n = 38 with six runs. The number of runs did not differ between TSST (5.22 ± 0.16 runs; mean ± SEM) and f-TSST participants (5.53 ± 1.43 runs; $t(62) = 0.10$, p = 0.322). Preprocessing of fMRI data included slice time correction and spatial realignment.

For MVPA analyses in native space, we did not spatially smooth the data. We specified a single-item general linear model (GLM):[9] Every item occurred twice in each run and was modeled as one predictor, resulting in 76 predictors for each of the six experimental runs. In addition, we modeled the oddball trials with a fly as one predictor and the break in the middle of each run as another predictor. Inter-trial intervals (where the fixation cross was presented) were not explicitly modeled. All regressors were convolved with the hemodynamic response function before entering the GLM. We did not model temporal derivatives. We included residual head motion as estimated in the realignment preprocessing step and slow artifactual trends as nuisance regressors.[9] β-files resulting from the GLM estimation were divided by the residual variance (ResMS) to ensure that voxels with high noise would not exert an overly strong influence on the results. We thus obtained maps of *t*-values for the individual objects in every run. Finally, we extracted the voxels of the respective ROIs (left/right amygdala and left/right hippocampus) in native space and performed the RSA. For loading the nifti-files into MATLAB, we used the "Tools for NIfTI and ANALYZE image" toolbox.[69]

For the RSA, we performed a Spearman correlation of the voxel values of each condition, i.e., each stimulus, with every other condition. Correlation values were Fisher-Z-transformed to improve normality. Next, we excluded correlations of conditions within runs to reduce the influence of autocorrelations. Finally, we averaged across the repetitions of every stimulus across runs resulting in a 76 by 76 matrix containing the similarities between all objects and faces for each participant and each ROI. Higher correlations indicate that the voxel patterns are more similar to one another, i.e., that their distance in representational space is lower. In other words, the similarity between voxel patterns is conceptualized as proximity in representational space. Please note that negative correlation values in the context of RSA do not speak for a negative association between voxel patterns, but instead indicate low similarity (e.g., Dimsdale-Zucker and Ranganath[70]). In the following steps, we extracted similarities from these matrices to calculate measures of representational structure.

We extracted four measures to describe the representational structure:

(1) Object-object similarity: We defined object-object similarity as the object-specific similarity of a central or of a peripheral object with other central or peripheral objects, respectively (Figure 3A).
(2) Object-distractor similarity: We defined object-distractor similarity as the object-specific similarity of an object with the distractors (Figure 3A).
(3) Similarity difference score: We defined the similarity difference score as the object-specific difference between object-object similarity and object-distractor similarity (Figure 3A). This measure reflects how much more similar the objects are among themselves as compared to the distractors. In other words, it indicates how much stronger the features of the episode are bound compared to features that have not been part of the episode. The similarity difference score describes a crucial aspect of the representational similarity structure: By itself, a high similarity among objects from the stressful episode does not indicate whether these objects also become distinct from other, distractor objects, which is necessary to remember that these objects were indeed part of the stressful episode. Thus, the relative difference between "object-object similarity" and "object-distractor similarity" allows distinguishing objects from distractors. Specifically, objects may either become more dissimilar from one another compared to the distractor objects (distinctiveness hypothesis), or they may become more similar to one another compared to the distractor objects (generalization hypothesis).
(4) Stressor-object similarity: We defined stressor-object similarity as the object-specific similarity of the central or peripheral objects with the stressor, i.e., the face of the committee member. For central items, we selected only the similarities with the face of the committee member who manipulated the object during the stress procedure (Figure S1A). Peripheral objects were per definition not paired to a committee member face, so we used the similarities to both faces (Figure 4A).

In addition, we extracted two covariates from the data:

(1) Representational reliability: We defined object-specific representational reliability as the mean similarity of each object with itself across the functional runs. This measure reflects how much information about the object is contained in the voxel pattern. However, it does not relate to the representational structure, as it does not contain any information about the relationship between objects. Representational reliability for easy distractors and faces did not differ between experimental conditions (both $F \leq 0.42$, both $P_{Bonferroni} \approx 1$). Likewise, representational reliability for difficult distractors did not differ between experimental conditions (TSST versus f-TSST), object type (central versus peripheral) and their interaction (all $F \leq 1.82$, both $P_{Bonferroni} \geq 0.177$). Only for objects, we encountered a significant main effect of object type ($F(1,1040) = 17.07$, $P_{Bonferroni} < 0.001$) with higher representational reliability for central objects, but no effects of stress ($F(1,1040) = 0.64$, p = 0.426) or their interaction

Please cite this article in press as: Bierbrauer et al., The memory trace of a stressful episode, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.09.044



($F(1,1040) = 0.21$, $p = 0.644$). Representational reliability thus seems to be affected by the saliency of the object representation. Consequently, we decided to include it as a covariate in all analyses including fMRI measures of representational structure (results were similar without this covariate, Table S2).

(2) Mean activity (univariate): Using correlations as a measure of similarity normalizes for overall activation and variability of activity across space and is thus not sensitive to global activity differences.[9] However, in order to make sure that univariate BOLD activity differences would not affect the RSA results, we ran control analyses where we added the object-specific mean univariate activation across the ROI for each item as a covariate. We found a marginally significant main effect of stress ($F(1,62.0) = 3.89$, $p = 0.053$), indicating that stressed participants have a tendency for higher amygdala activity. There was no difference between central and peripheral objects ($F(1,1094.3) = 0.70$, $p = 0.404$) and no interaction effect ($F(1,1094.3) = 0.03$, $p = 0.867$). Due to the marginally significant main effect of stress, we decided to perform a control analysis including mean activity as an additional covariate (results were similar with this covariate, Table S2).

## Statistical analyses

A priori comparisons between experimental groups were assessed using two-sample t tests (Table S1). *P*-values were Bonferroni-corrected for the number of tests, i.e., 10 comparisons on day 1 and 4 comparisons on day 2.

We calculated subject-specific descriptive means for plots by averaging memory performance or representational measures across all objects of each object type (central or peripheral) separately for each participant.

We fitted mixed linear models using the lme4-package[63] in R. *P*-values were calculated using the lmerTest-package.[64] We used type III sum of squares such that the main effects/lower order terms were controlled for interactions/higher order terms. For all models, we used qualitative analysis of Q-Q-plots to ensure normality of model residuals. If normality was violated, we either transformed the data (cortisol), excluded outliers (cortisol, spatial memory, representational measures, and covariates), or used a permutation test to derive *P*-values (free recall).

The exclusion of outliers was performed separately for each analysis. We calculated mean scores for central and for peripheral objects for every continuous predictor or outcome variable. We excluded participants with values more than three scaled median absolute deviations (MAD) away from the median from the respective analysis as this procedure is more robust than other "classical" ways of outlier detection.[71] There were no outliers for free recall performance (number of recalled objects), recognition memory performance (Pr and memory strength); 2 outliers for spatial memory performance (participant IDs: 14, 65); 2 outliers for cortisol increase (Baseline to +10 min; IDs: 30, 44); 6 outliers for the covariate representational reliability (ID's: 4, 10, 13, 34, 44, 55); no outliers for the covariate mean activity; 4 outliers for the similarity difference score (ID's: 4, 13, 20, 33); 4 outliers for object-object similarity (ID's: 4, 13, 33, 44); 3 outliers for object-distractor similarity (ID's: 13, 20, 44), and 3 outliers for stressor-object similarity (ID's: 7, 41, 58).

Additionally, we made sure that variance inflation factor (VIF) would not exceed a value of 5 indicating multicollinearity (all VIFs $\leq$ 1.03; VIFs estimated for models excluding interaction terms).

Generally, we applied a threshold of $\alpha = 0.05$ (two-tailed) for assessing statistical significance and we corrected for multiple comparisons using Bonferroni correction; details are described with the respective models. For significant effects, we report Cohen's $f^2$, a measure of local effect size that has been suggested for mixed linear models.[72–74] It reflects the proportion of variance that the respective predictor uniquely accounts for.[72,73] We used the effectsize-package[75] in R to calculate Cohen's $f^2$.

We log-transformed cortisol response values to ensure normality. Log-transformed cortisol response and subjective stress response (mean PA, mean NA, and mean SSES scores) were analyzed using mixed linear models with "time" [cortisol response: "baseline" (BL), "1 min after (f-)TSST" (+1), "10 min after (f-)TSST" (+10), "25 min after (f-)TSST" (+25); subjective stress response: "before" versus "after" (f-)TSST] as within-subject factor, "experimental condition" (TSST versus f-TSST) as between-subject factor, and "subject" as random factor (Figure 2A). *P*-values of the follow-up simple effects tests were Bonferroni corrected for the number of comparisons (i.e., 4 times, 2 times, or 2 experimental conditions).

Free recall (number of recalled objects), recognition memory (Pr), and spatial memory performance (distance error) averaged across object categories (i.e., central and peripheral) were analyzed using mixed linear models with "object type" (central versus peripheral) as within-subject factor, experimental condition (TSST versus f-TSST) as between-subject factor, and subject as random factor (Figures 2C, S2A, and S2B). *P*-values of the follow-up simple effects tests were Bonferroni corrected for the number of comparisons (i.e., 2 object types or 2 experimental conditions).

Free recall performance for central and peripheral objects was not normally distributed (as indicated by the Q-Q-plots of the model residuals; see also Kolmogorov-Smirnov-test: both $D(64) \geq 0.452$, both $p < 0.001$) and it was not possible to exclude outliers or transform the data such that data would follow normal distribution. We thus report *P*-values of a permutation test: We either shuffled the experimental condition (i.e., whether participants were part of the TSST group or of the f-TSST group) to assess the stress and interaction effect or the object type (i.e., whether an object was central or peripheral). We performed the mixed linear model on shuffled data 10.000 times and calculated the *P*-values of the *F*-tests and of the post hoc simple effects t tests by assessing the quantile of the actual *F*- and *t*-values with respect to the random distributions of *F*- or *t*-values. In the description of the results, we report the parametric *dfs*, *F*- and *t*-values together with the *P*-values of the permutation tests.

In order to characterize the recognition memory comprehensively, we predicted the AUC for ROC curves from "object type" (central versus peripheral), "distractor type" (easy versus difficult) and "experimental condition" (TSST versus f-TSST; see Figure S2D; Table S3). *P*-values of the follow-up simple effects tests were Bonferroni corrected for the number of comparisons (i.e., 2 object

Please cite this article in press as: Bierbrauer et al., The memory trace of a stressful episode, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.09.044



types, 2 distractor types, or 2 experimental conditions). Additionally, we performed t tests to compare hit rates and false alarm rates of central and peripheral objects as well as easy and difficult distractors between TSST and f-TSST (see Figure S2E; Table S3).

To analyze if there were *a priori* differences of representational measures between the object groups, we used the fMRI data from the control experiment and from the main experiment. For the control experiment, we extracted the measures of representational structure in the same way as for the main experiment. For both datasets, we compared representational reliability, the similarity difference score, object-object similarity, and object-distractor similarity for the four ROIs across the two object groups (episode A and B, see Figure S1) using paired sample t tests. *P*-values of the t tests were Bonferroni corrected for 4 ROIs (i.e., left/right amygdala and left/right hippocampus).

Measures of representational structure, i.e., the similarity difference score, object-object similarity, object-distractor similarity, and stressor-object similarity were analyzed on the level of individual objects using mixed linear models with object type (central versus peripheral) as within-subject factor, experimental condition (TSST versus f-TSST) as between-subject factor, object-specific "representational reliability" as covariate, and subject as random factor. *P*-values of the omnibus tests were Bonferroni corrected for 4 ROIs (i.e., left/right amygdala and left/right hippocampus). *P*-values of the follow-up simple effects tests were Bonferroni corrected for the number of comparisons (i.e., 2 object types or 2 experimental conditions). As a follow-up analysis, we performed the models on the similarity difference score with easy and difficult distractors separately.

In two additional control analyses, also object-specific "univariate mean activity" was added as covariate and models were re-run without any covariates (Table S2). We also employed the same models to analyze the relationship between experimental condition, object type, and the covariates "representational reliability" and "mean activity in the left amygdala" (for the results of these control analyses, see above "FMRI analyses," section about covariates).

To test how measures of representational structure in the left amygdala relate to memory performance, we created mixed linear models on the level of individual objects to predict "memory strength." We defined memory strength as the object-specific rating minus the participant-specific mean rating of all objects in order to correct for participant-specific response bias. We inverted the values (simply by subtracting them from 7), such that higher values reflected higher memory strength. We included object type (central versus peripheral) as within-subject factor, experimental condition (TSST versus f-TSST) as between-subject factor, and added either the object specific "similarity difference score" or "stressor-object similarity" in the left amygdala as a parametric predictor. We included object-specific representational reliability in the left amygdala as covariate and subject as random factor. *P*-values of the follow-up simple effects tests were Bonferroni corrected for the number of comparisons (i.e., 2 object types or 2 experimental conditions).

We predicted the similarity difference score using stressor-object similarity and object type (central versus peripheral) as within-subject factors and experimental condition (TSST versus f-TSST) as between-subject factor. We included object-specific representational reliability in the left amygdala as covariate and subject as random factor. The rationale behind this analysis is that higher representational similarity (i.e., lower representational distance) between stressor and objects might reflect a clustering of objects around the stressor. This clustering could at least partially explain higher similarity difference scores.

Next, we predicted memory strength by object type (central versus peripheral) as within-subject factor, experimental condition (TSST versus f-TSST) as between-subject factor, and included both, the object specific similarity difference score and stressor-object similarity in the left amygdala as within-subject factors. We included object-specific representational reliability in the left amygdala as covariate and subject as random factor. This control analysis was performed to see how the two factors interact with each other in predicting memory.

As a follow-up analysis, we selected only central items for participants in the TSST condition and predicted memory strength (again on the level of individual objects) using "stressor-object similarity" in the left amygdala as within-subject factor and "cortisol increase" as between-subject factor. Cortisol increase was defined as the increase in saliva cortisol from baseline to minute 10 after the stress procedure (as this was the peak in the cortisol curve, see Figure 2A). We included "object-specific representational reliability" in the left amygdala as covariate and "subject" as random factor. There was a trend for an interaction between cortisol increase and stressor-object similarity ($F(1,237.2) = 3.42$, $p = 0.066$).

For the post hoc tests, we discretized cortisol increase into responders (cortisol increase $\geq$ 1.5 nmol/l, n = 19, mean ± SD: 15.99 ± 7.74 nmol/l) and non-responders (n = 9, mean ± SD: −2.91 ± 1.98 nmol/l). The criterion of 1.5 nmol/l is well-established.[38] We estimated the strength of the relationship between stressor-object similarity and memory strength for the two groups based on the model. *P*-values of the follow-up simple effects tests were Bonferroni corrected for the number of comparisons (i.e., 2 groups: responders and non-responders). In Figure 4E, we show the model estimates for responders and non-responders as well as for single participants.

In an additional analysis, we tested whether the effects that we found for all objects could also be found in the subcategory of "remembered" objects. Therefore, we applied the same models (predicting the similarity difference score, object-object similarity, object-distractor similarity, and stressor-object similarity) only to the remembered objects (rating < 4; see Figures S4A–S4D; Table S4).

In order to compare the structure of neural representations between left and right amygdala, we modeled the similarity difference score and object-object similarity as a function of "experimental condition" (TSST versus f-TSST), "object type" (central versus peripheral), and "hemisphere" (left versus right; see Figures S4E–S4G; Table S4). We included "subject" as a random factor and "representational reliability" as a covariate. We also re-ran the model to predict memory strength from the stressor-object similarity with "experimental condition," "object type," and "hemisphere" as regressors. Post hoc tests were performed to reveal differences between left and right amygdala.