Suppression weakens unwanted memories via a sustained reduction of neural reactivation

Ann-Kristin Meyer and Roland G. Benoit

Max Planck Institute for Human Cognitive and Brain Sciences, Max Planck Research Group: Adaptive Memory, Leipzig, Germany

Aversive events often turn into intrusive memories. However, prior evidence indicates that these memories can be forgotten via a mechanism of retrieval suppression. Here, we test the hypothesis that suppression weakens memories by deteriorating their neural representations. This deterioration, in turn, would hinder their subsequent reactivation and thus impoverish the vividness with which they can be recalled. In an fMRI study, participants repeatedly suppressed memories of aversive scenes. As predicted, this process rendered the memories less vivid. Using a pattern classifier, we observed that it did diminish the reactivation of scene information both globally across the grey matter and locally in the parahippocampal cortices. Moreover, in the right parahippocampal cortex, a stronger decline in vividness was associated with a greater reduction in generic reactivation of scene information and in the specific reinstatement of unique memory representations. These results support the hypothesis that suppression deteriorates memories by compromising their neural representations.

memory | voluntary forgetting | suppression | reinstatement

✉ ***annmeyer@cbs.mpg.de, rbenoit@cbs.mpg.de***

Introduction

Memories of the past are not always welcome. There are experiences that we would rather not think about, yet that involuntarily intrude into our awareness. Research over the last two decades has demonstrated that we are not at the mercy of such unwanted memories: we can control them by actively suppressing their retrieval (1–3). This process weakens the memory and can eventually cause forgetting (4). We here seek to tie the phenomenological weakening of a suppressed memory to its neural basis.

Neuroimaging research has made strides in determining the mechanisms that prevent unwanted retrieval. It has consistently shown that retrieval suppression is mediated by an engagement of the right dorsolateral prefrontal cortex (dlPFC) that, in turn, reduces activity in the hippocampus (5–13). This pattern has been interpreted as reflecting a top-down inhibition of critical hippocampal retrieval processes (5–7). The dlPFC may exert such control either via a gating of entorhinal input into the hippocampus or by modulating hippocampal activity via the thalamic reuniens nucleus (14).

Retrieval suppression, however, does not only entail a downregulation of the hippocampus. For example, when the unwanted memories comprise images of complex scenes, it is also accompanied by reduced activation of the parahippocampal cortex (PhC) (8). This region particularly supports memories for scenes (15–18) and its activity during retrieval scales with the detailedness (19, 20) and vividness (21–23) with which they can be remembered. Similarly, suppressing memories of objects leads to reduced activation in object-sensitive parts of the inferior temporal cortex (10, 13). Retrieval suppression thus also entails a reduction of activity in cortical regions that encode the particular content of the suppressed memory (see also (6, 9–11).

Successful memory retrieval is based on the reinstatement of the neuronal activity pattern that was present during initial encoding (24–26). The local reductions in brain activity during suppression may accordingly prevent unwanted retrieval by hindering such reinstatement ((10); see also (27)). In humans, memory reinstatement has been demonstrated using multivariate analyses of fMRI data. These approaches exploit the distributed pattern of activity across voxels as a proxy of a memory's neural representation. Evidence for reinstatement has thus been observed across distributed cortical and subcortical areas (26, 28–30). Specifically, for the successful retrieval of scenes, it has also been shown locally in the PhC (16, 31, 32). In turn, there is some evidence that concerted attempts to suppress an unwanted memory indeed *prevent* reinstatement at that time ((10, 33, 34); see also (35)).

Despite our evolved understanding of the mechanisms mediating suppression, there is little evidence for the critical neural after-effects: Why do previously suppressed memories remain difficult to recall? Here, we test the hypothesis that suppression deteriorates the memory's neural representation (5, 36). It would thus prevent later reinstatement even when people then try to intentionally recall that memory (see also (27)). A deficient reinstatement, in turn, would affect the vividness with which the memory can be recalled.

To test this hypothesis, we conducted an fMRI study using an adapted *Think/No-Think* procedure (37, 38). First, participants learned associations between neutral objects (cues) and aversive scenes (target memories) Fig.1a. During the suppression phase, they were then scanned by fMRI while they again encountered the cues. The participants were repeatedly prompted to recall the associated target for some of the cues (*recall* condition), whereas they were requested to prevent the retrieval of the targets for other cues (*suppress* condition). In short, we instructed participants to remain focused on the cue while trying to block out all thoughts of the accompanying target memory without engaging in any distracting activity (7, 39). Importantly, we did not present a third of the cues during this phase (*baseline* condition). These cues and their associated targets thus serve as a baseline for the fading of memories that simply occurs as a passage of time (i.e., without any suppression attempts).

Critically, we also had participants recall each target in response to its cue both before and after the suppression phase. During these pre- and post-tests, they also reported the vividness with which they could recall the memories. We thus assessed the phenomenological quality of the memories at the same time that we probed their neural reinstatement. Finally, participants engaged in a separate task that allowed us to train a pattern classifier to detect neural reactivation of complex and aversive scenes.

We tested our hypothesis by tracking the impact of suppression on neural reinstatement. First, we expected that suppression would be associated with reduced scene reactivation (10, 33, 35), both distributed across the brain and more regionally specific in the PhC. We critically predicted that this effect would also linger on - as indexed by lower post-test reactivation of previously suppressed memories. Second, in addition to a reduced reactivation of general scene information, we also predicted a weaker parahippocampal reinstatement of activity patterns that are unique to individual scene memories. To the degree that weaker reinstatement constitutes the basis for the reductions in vividness, we finally expected a relationship between these phenomenological and neural effects.

Results

Preventing retrieval yields the typical pattern associated with memory suppression. We first sought to establish whether our procedure elicited the activation pattern that has consistently been associated with retrieval suppression (e.g. (7, 13, 40, 41)). Suppressing versus recalling an aversive scene indeed led to increased activation in a number of brain regions including the right dlPFC and reduced activation in, amongst others, the bilateral hippocampi and PhC (Fig.1b, Supplemental Tables 1 & 2) This pattern suggests that our procedure successfully induced a mechanism of retrieval suppression.

In the following, we test the hypothesis that this mechanism impairs subsequent retrieval attempts by hindering reinstatement of the neural memory representation. We thus examine suppression-induced changes in the phenomenological quality of the memories and their neural basis. These analyses focus on the critical comparison of the *suppress* versus *baseline* conditions. (In the supplement, we explore possible effects of retrieval practice (42–44), i.e., contrasts of the *baseline* versus *recall* conditions.)

Suppression renders memories less vivid. We assessed the impact of suppression on the phenomenological quality of the memories by examining their change in vividness from the pre-test to the post-test. Indeed, there was a greater reduction for *suppress* than *baseline* memories as indicated by a significant interaction between time of test (pre, post) and condition (*baseline*, *suppress*) ($F(1, 32) = 46.18$, $p < .001$, η^2_G (generalized eta squared) = .034). However, the main effects of time of test ($F(1,32) = 28.87$, $p < .001$, η^2_G =.063) and condition ($F(1,32) = 4.22$, $p = .048$, $\eta^2_G = .007$) were also significant. Follow-up tests showed that suppression impoverished the vividness of the memories ($t(50.9) = 8.04$, $p < .0001$), whereas *baseline* memories were not significantly affected by the mere passage of time ($t(50.9) = 1.29$, $p = .20$) (Fig.1c).

We obtained largely the same pattern in a behavioral study with an independent sample: again, the critical interaction between time of test and condition was significant ($F(1, 28) = 8.85$, $p = .006$, $\eta^2_G = .015$) (in addition to the main effects of time of test, $F(1, 28) = 21.78$, $p < .001$, $\eta^2_G = .101$, and of condition, $F(1, 28) = .11$, $p = .008$, $\eta^2_G = .02$). *Baseline* and *suppress* memories did not differ on the pre-test ($t(53.8)$= .32, $p = .75$), though they did on the post-test following the suppression phase ($t(53.8)$= 4.07, $p = .0002$). However, the follow-up tests showed a reduction in vividness for suppressed ($t(44.4) = 5.53$, $p < .0001$) and also for *baseline* memories ($t(44.4) = 2.59$, $p = .0001$).

Consistent with prior research (4), suppression thus had a detrimental, replicable impact on people's ability to recall the avoided memories. Importantly, we assessed the phenomenological quality of the memories during exactly those retrieval attempts that also provide the basis for our critical fMRI analyses. That is, in the following, we examine not only whether there is less reactivation of a memory during suppression (10, 33), but also the hypothesis that this effect then lingers on during these subsequent recall attempts.

Training the linear classifier to detect scene reactivation. Memory retrieval reactivates the perceptual and conceptual representations elicited during encoding (45, 46). To quantify the degree of such reactivation on a given trial, we trained a linear support vector machine on independent data (47). Specifically, the classifier learned to distinguish brain states associated with the perception of intact aversive scenes (similar to the ones used in the main task) versus morphed versions of the scenes. The morphed scenes were created via a diffeomorphic transformation that renders them unrecognizable while preserving their basic perceptual properties. Compared to conventional methods, such as scrambling, morphing has been shown to elicit neural activation that is more similar to activation induced by intact images (48).

Given the widespread nature of memory representations (25, 28, 29), we sought to test for global reactivation by training a classifier on all voxels of the respective participant's grey matter mask. Using cross-validation, the classifier reached a mean accuracy of 80.3% (*SD* = 17.4) on the training data, corroborating that it was able to distinguish brain states associated with the presentation of intact versus morphed aversive scenes ($t(32) = 10$, $p < .001$). We thus were

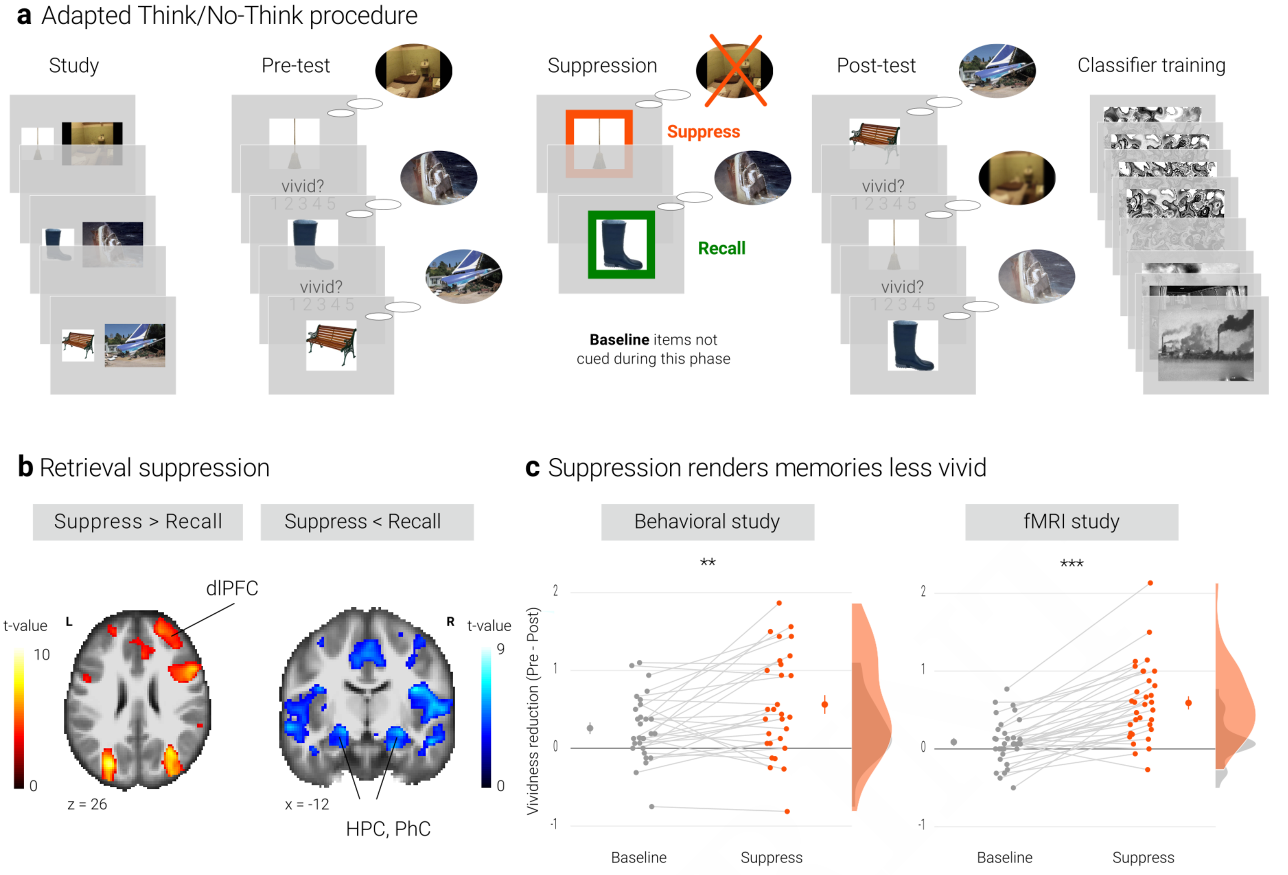

Fig. 1. Experimental procedure, univariate MRI and behavioral results. a) Illustration of the adapted Think/No-Think procedure. Participants studied 48 critical associations between unique objects and aversive scenes. Both during a pre- and a post-test, they covertly recalled all the scenes in response to the objects and rated the vividness of their recollection. In between these two tests, they performed the suppression phase. Specifically, for objects presented in a green frame, participants repeatedly recalled the associated scene (*recall* condition). By contrast, for objects presented in a red frame, they suppressed the retrieval of the associated scene (*suppress* condition). Note that we did not present a third of the objects during this phase (*baseline* condition). Finally, participants performed a 1-back task that served to train a pattern classifier in detecting evidence for scene reactivation. **NB** As per bioRxiv policy, for the figure, we replaced some scenes featuring people by stand-ins. In the original stimulus set, each object cue also features in its paired scene. **b)** The suppression phase yielded the typical activity pattern associated with retrieval suppression, including greater activation in the right dorsolateral prefrontal cortex (dlPFC) and reduced activation in the hippocampus (HPC) and parahippocampal cortex (PhC) during *suppress* versus *recall* trials. For display purposes, the images are thresholded at $p < .001$,uncorrected , with a minimum cluster size of 50 voxels. **c)** Suppression caused a reduction in vividness from the pre- to post-test that exceeded any change due to the passage of time as indexed by the *baseline* condition. This effect replicated across the fMRI study (n = 33) and a behavioral study (n = 30) with an independent sample. Large dots indicate the mean; error bars indicate standard error of the mean. *** $p < .001$, ** $p < .01$.

able to use the classifier to estimate the reactivation of scene information on a given trial by calculating the dot product of the trial's activation map and the classifier's weight pattern (49, 50).

We also sought to test for more localized reactivation of scene information in the PhC, given the preferential engagement of this region for scene memories (15, 51). Towards this end, we manually traced the posterior parahippocampal cortices on each individual anatomical scan (52, 53) (Fig.2) and trained classifiers separately for the masks from the left and right hemisphere. These classifiers reached average cross-validation accuracies of 77.3% (SD = 16.6; $t(32) = 9.5$, $p <$ .001) and 82.6% (SD = 16.6; $t(32) = 11.3$, $p < .001$), respectively.

We further validated our approach by examining the correspondence between these reactivation scores and the vividness with which the memories could be recalled. This analysis was based on the pre-test, given that memories at that stage are still unconfounded by possible effects of the subsequent experimental manipulations. Across all trials of all participants, greater scene reactivation was indeed associated with more vivid recollections in left (Spearman's rho = .07, p = .008) though not right PhC (Spearman's rho = .03 , p = .21).

Reduced scene reactivation *during* suppression. The previous section established that the classifier provides a measure for the reactivation of scene information. Before turning to the after-effects of suppression, we first examined whether there is less evidence for scene reactivation while participants intentionally try to suppress rather than to recall a memory. This analysis did yield evidence for reduced scene reactivation globally ($t(32) = -7.04$, $p < .001$, $d = -1.22$) as well as locally in the left ($t(32) = -2.84$, $p = .008$, $d = -0.5$), though not right PhC ($t(32) = 0.6$, $p = .55$, $d = 0.10$) (Fig.2b).

Importantly, these data indicate that participants were successful at controlling the retrieval of unwanted memories. At the same time, they further validate the use of the classifier as a measure of memory reactivation.

Reduced global scene reactivation *following* suppression. Suppressed scenes were recalled less vividly than baseline scenes. We had hypothesized that this suppression-induced decline of the memories reflects a distortion of their neural representations. Such distortion, in turn, would hinder their reactivation on subsequent retrieval attempts. We thus expected reactivation scores for *suppress* memories to decline from the pre-test to the post-test to a degree that exceeds a possible decline for *baseline* memories.

We tested for this effect by conducting an ANOVA on the global reactivation scores with the factors time of test (pre, post) and condition (*suppress*, *baseline*). This analysis yielded the expected significant interaction ($F(1{,}32) = 5.14$, $p = .03$, $\eta^2_G = .006$), reflecting diminished scene reactivation for suppressed ($t(54.2) = 2.23$, $p = .03$) but not for baseline memories ($t(54{,}2) = -0.2$, $p = .84$) (Fig.2c).

Reduced parahippocampal scene reactivation *following* suppression. The PhC is particularly involved in retrieving scene information (15–17) and, consistent with prior evidence (8), its activity was decreased during suppression. Accordingly, we had predicted that this region would also show less reactivation of scene information during the retrieval of previously suppressed memories.

This was corroborated by an ANOVA with the factors time of test (pre, post), condition (*baseline*, *suppress*), and hemisphere (left, right) that yielded the significant interaction between time and condition ($F(1{,}32) = 4.33$, $p = .046$, $\eta^2_G =$ .003) (in addition to a main effect of time, $F(1{,}32) = 8.83$, $p = .006$, $\eta^2_G = .17$). This effect reflected the expected reduction in scene reactivation for *suppress* ($t(53.3) = 3.6$, $p = .0007$) but not for *baseline* memories ($t(53{,}3)= 1.44$, $p = 0.16$) (Fig.2c).

A link between suppression-induced reductions in scene reactivation and vividness. Activity in the PhC has previously been associated with the number of details (19, 20) and the vividness (21–23) with which scenes can be recalled. We similarly observed that the recall of more vivid memories is accompanied by greater evidence for scene reactivation.

We accordingly hypothesized that a greater suppression-induced reduction in scene reactivation would lead to a greater reduction in vividness. We tested this account by exploiting the natural variation in people's ability to control unwanted memories: individuals showing a greater suppression-induced reduction in vividness should also yield reduced evidence for scene reactivation.

To test this account, we calculated, for each participant, a behavioral suppression-induced reduction score. This score was calculated as the reduction in vividness from the pre- to the post-test for suppressed memories, corrected for the reduction in vividness for baseline scenes:

$$\text{suppression-induced reduction} = (\text{pre}_{\text{suppress}}\text{-post}_{\text{suppress}}) - (\text{pre}_{\text{baseline}}\text{-post}_{\text{baseline}})$$

We thus obtained an index of the deterioration in vividness that exceeds any effects that simply occur due to the passage of time (7, 40).

Analogously, we calculated the degree of suppression-induced reductions in scene reactivation by subtracting the change score of the baseline memories from the score of the suppressed memories. If the reduction in reactivation is linked to the reduction in vividness, we expected a positive correlation between the behavioral and neural suppression-induced reduction scores. Indeed, using robust skipped Spearman's correlations, we found a significant effect for the right ($r = .46$, 95%-CI = [.08 .76]) and a trend for the left PhC ($r = .34$, 95%-CI = [-.05 .67], Fig.3).

Taken together, suppression led to a reduction of scene information on a global and local level. Moreover, the degree of reduced reactivation of scene information in the right PhC was linked to the decline of the memories' vividness. These data suggest that reduced reactivation reflects the failure to retrieve scene features that would have made the recollections more vivid. In the following, we examine this account more directly by assessing changes in the neural reinstatement of individual neural memory representations.

Suppression success is associated with weaker parahippocampal pattern reinstatement. The classifier results indicate that suppression hinders the subsequent reactivation of scene information. However, they do not address the question whether this effect reflects reduced reinstatement of information that is specific to a particular memory. In a next step, we thus used Representational Similarity Analysis (RSA) (54, 55) to examine the reinstatement of activity patterns that are unique to the individual memories. We focus this analysis on the PhC, where the neural reinstatement of a particular memory should yield a unique and replicable activity pattern (16, 31, 32). Specifically, we expected a similar activity pattern to emerge whenever participants recall the same scene memory.

We quantified similarity by computing the Pearson correlation (54, 55) of the activity patterns across the pre- and post-test. We then compared the similarity of a memory with itself (same-item similarity) and the similarity of a memory with all other memories of the same (e.g., *baseline*) condition (different-item similarity) (31, 56, 57). For the *baseline* memories, an ANOVA with the factors scene identity (same, different) and hemisphere (left, right) indeed yielded greater same- than different-item similarity ($F(1{,}32) = 10.59$, $p =$

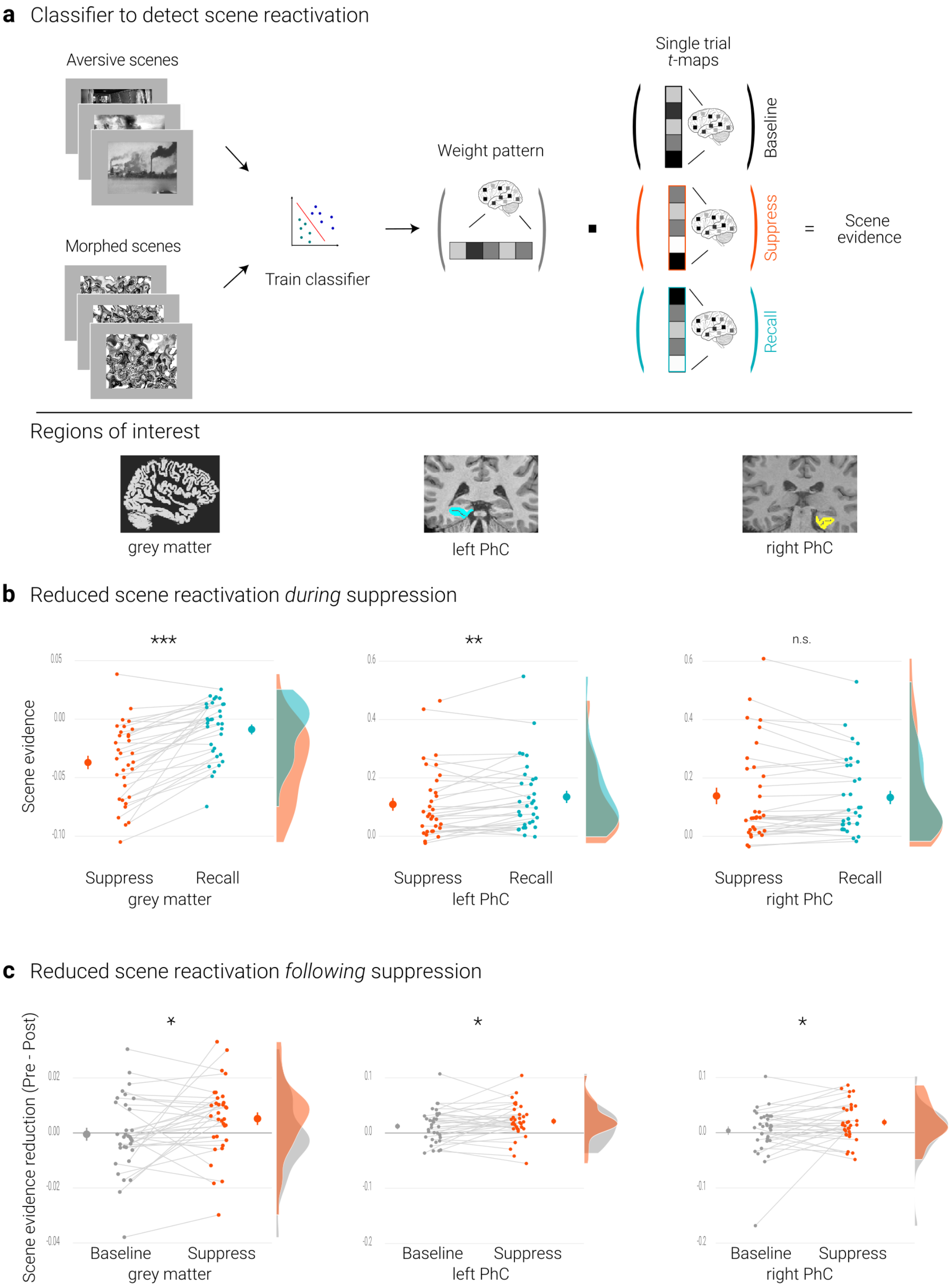

Fig. 2. Effects of suppression on scene reactivation. a) A linear support vector machine was trained on independent data to discriminate neural activity patterns associated with the perception of intact versus morphed aversive scenes. The dot product of the resulting weight pattern and single-trial *t*-maps was used as a proxy for scene evidence. We compared the scene evidence between conditions globally across the grey matter and more locally in the left and right parahippocampal cortices (PhC, manually traced on the individual structural images). **b)** Across the grey matter and locally in left PhC, there is less scene evidence while participants suppress than recall scene memories. **c)** The suppression-induced reduction in scene evidence lingers on after suppression: scene evidence decreases from the pre- to the post- test for suppressed memories but not for baseline memories. This was the case across the grey matter and in the PhC. Larger dots indicate the mean, error bars indicate standard error of the mean. *** $p <$ .001, ** $p <$ .01, * $p <$ .05, n = 33.

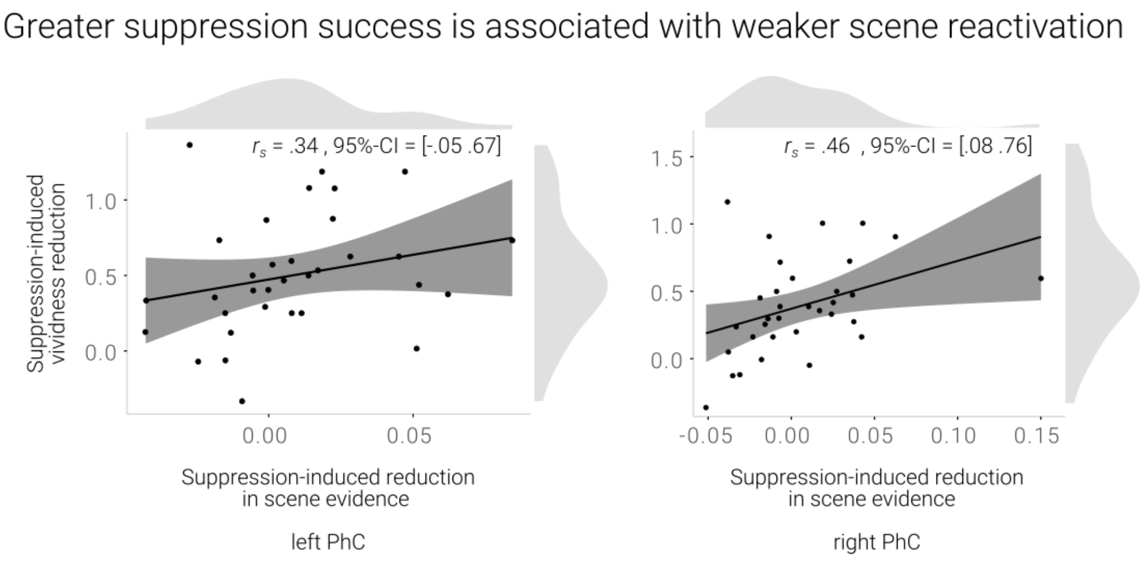

Fig. 3. a,b) A greater below-baseline reduction in vividness is associated with a greater below-baseline reduction in scene evidence in right PhC as indicated by a robust skipped Spearman's correlation. The left PhC showed a trend only for this effect. Black lines indicate linear regression lines, dark grey shades indicate 95% - confidence intervals, PhC: parahippocampal cortex.

.003, η^2_G = .006) (Fig.4b). The data thus provide evidence for the replicable reinstatement of neural representations that are unique to the individual memories.

By contrast, for the *suppress* memories, there was only a trend for a numerically smaller effect (F(1,32)= 3.38, p = 0.075, η^2_G = .004) (in addition to a significant main effect of hemisphere, $F(1{,}32) = 6.68$, $p = 0.015$, η^2_G = .031). However, a combined analysis with the added factor condition (*baseline*, *suppress*) yielded the main effect of hemisphere (F(1,32 = 5.41, p = .027, η^2_G = .022) but not the critical interaction between scene-identity and condition (F(1,32) = 0.46, p = .501). Though there was only weak evidence for the reinstatement of suppressed memories, the effect was -overall- not significantly smaller than for the baseline memories.

However, we had hypothesized that individuals who were more successful at suppression (as indicated by a greater reduction in vividness) should show evidence for a greater decline in neural reinstatement. As for the reactivation scores above, we thus examined the association between the behavioral suppression-induced reduction scores and the difference in reinstatement for baseline versus suppressed memories. The latter was computed as

reinstatement = $r_{\text{same-item}}$ - $r_{\text{different-item}}$

suppression-induced reduction =
reinstatement$_{\text{baseline}}$ - reinstatement$_{\text{suppress}}$

Thus, a greater value indicates a greater reduction in memory specific reinstatement. Mirroring the results of the pattern classifier, the skipped Spearman's correlation between the behavioral and neural effects was significant in the right (r = .39 , 95%-CI= [.02 .68]) though not left PhC (r = .019, 95%-CI = [-.37 .41] (Fig.4c)).

Discussion

Research over the last two decades has demonstrated that we are not at the mercy of unwanted memories (37). Instead, we can intentionally suppress their retrieval, a process that weakens the avoided memory and eventually can lead to forgetting (4, 38). Though, we have made strides in understanding the mechanisms supporting retrieval suppression (5–13), there is little evidence for the neural consequences that underlie the suppression-induced changes in a memory's retrievability.

In this study, we sought to tie the phenomenological weakening of a memory caused by suppression to its neural basis. Successful episodic memory retrieval entails the reinstatement of a memory's representation (24, 26, 45, 58, 59). It is fostered by hippocampal processes that complete the neuronal pattern of the original experience (e.g., of a particular scene) from a partial pattern provided by an adequate retrieval cue (e.g., of an object that was also part of the scene) (60–62). This process leads to the cortical reinstatement of a memory across the regions that had been involved in its original encoding (18, 28, 63).

Reinstatement has been examined in humans using fMRI by exploiting the distributed pattern of activity across voxels as a proxy of a memory's neural representation. Successful retrieval (e.g., of a particular scene) has thus been shown to be accompanied by a reactivation of categorical information (e.g., scene information) that is both, widely distributed (26, 28–30) and localized to specific brain regions (31, 32).

In the current study, the degree of scene evidence on a given trial scaled with the vividness with which a memory could be retrieved. This extents prior evidence showing that activation, particularly in the PhC, is stronger when scenes are recollected more vividly and in greater detail (19–23). Our data

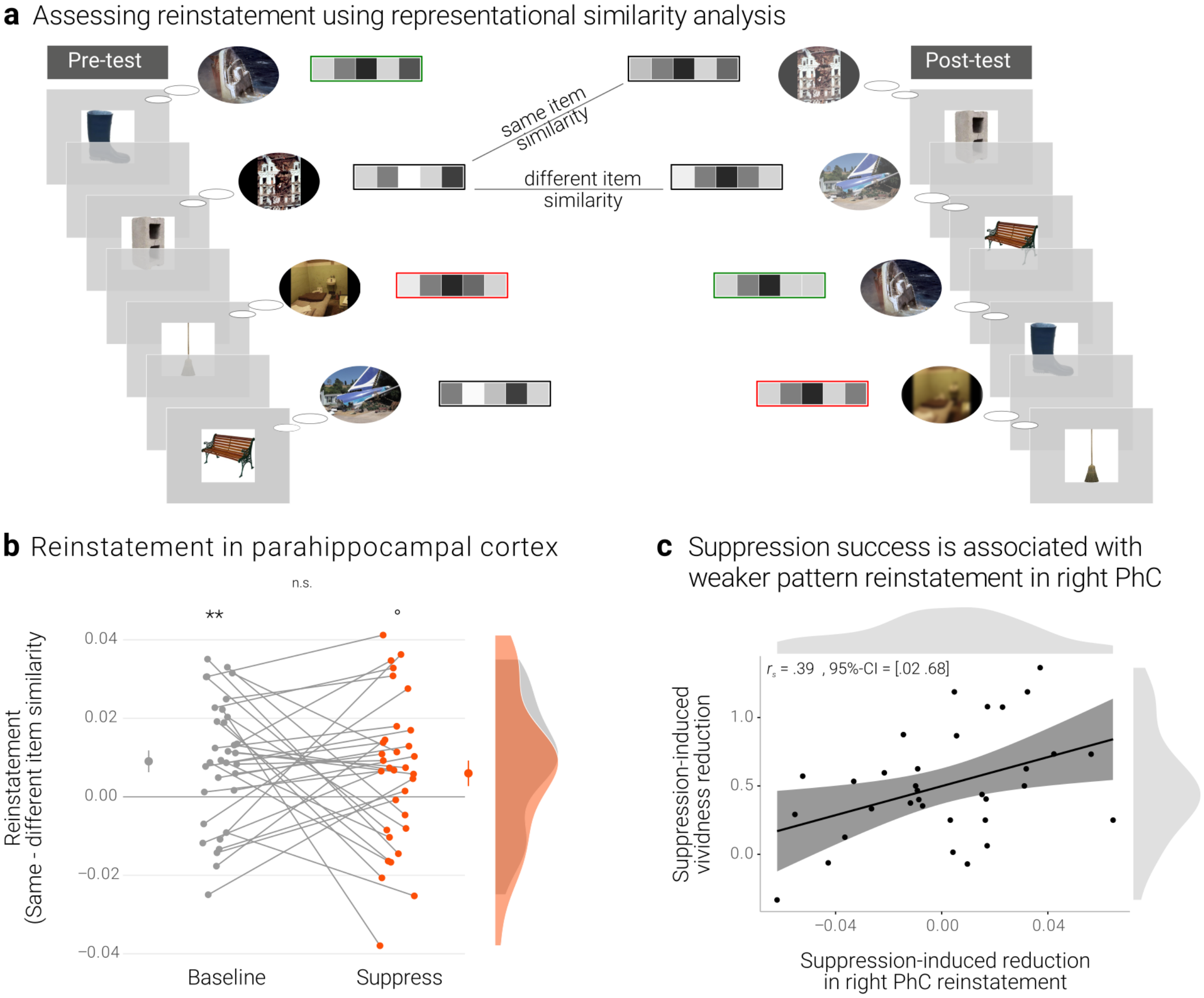

Fig. 4. Effects of suppression on individual memory representations. a) We estimated the reinstatement of the neural memory representations by assessing the similarity of the activity patterns across the pre- and the post-test. We take the difference between the same-item and different-item similarity as an index of neural reinstatement. **NB** As per bioRxiv policy, for the figure, we replaced some scenes featuring people by stand-ins. **b)** In the PhC, the difference between same- and different-item similarity was significant for the baseline memories. These memories were thus consistently reinstated across the two tests. By contrast, the suppressed memories only showed a trend for this effect, though the critical interaction between condition and time of test was not significant. Large dots indicate the mean, error bars indicate standard error of the mean. **c)** A greater suppression-induced reduction in vividness was associated with a greater suppression-induced reduction in reinstatement in the right PhC (as indicated by a robust skipped Spearman's correlation). Black lines indicate linear regression lines, dark grey shades indicate 95% - confidence intervals, PhC: parahippocampal cortex. ** $p < .01$, * $p < .05$, ° $p < .1$, n = 33

thus further validate the use of classifier evidence as a marker of memory reactivation. Consequentially, we expected that intentional attempts to prevent retrieval should lead to reduced scene reactivation. This was the case in the current study across the grey matter as well as for the left PhC (see also (10, 33, 34)). Participants thus seem to have successfully suppressed unwanted retrieval.

The analysis of the phenomenological ratings showed that suppression diminished the vividness with which the memories could subsequently be recalled. This finding adds to the extant literature by showing that suppression does not only affect the objective availability of a memory in an all-or-none fashion (4, 10, 13, 38). It rather also gradually diminishes the subjective quality of the information that can be retrieved (see also (64, 65)). Such graded forgetting can be beneficial, for example when it allows for the continued conscious access to an aversive past event while dampening its affective impact (2, 66).

Critically, we sought to test the hypothesis that the fading of a suppressed memory results from a deterioration of its neural representation. This deterioration, in turn, would manifest as impeded neural reinstatement on subsequent retrieval attempts. To test this account, we tracked changes in the reactivation of suppressed representations.

The classifier analysis yielded evidence for less scene reactivation during the recall of suppressed than of baseline memories. This was the case at a global level across the grey matter as well as locally in the PhC. Moreover, consistent with the contribution of the PhC to mnemonic vividness (21), we observed that people who displayed a greater reduction in scene reactivation in the PhC also experienced a greater suppression-induced reduction in vividness. Indeed, disruptions of cortical memory representations such as in the PhC may particularly lead to these graded effects of forgetting (10, 30, 65). That is, as cortical representations get progressively weakened, they may become increasingly susceptible to interference from overlapping representations of similar memories (67).

By comparison, at the hippocampal level, different representations are encoded in an orthogonal fashion due to pattern separation supported by the dentate gyrus (62, 68). Accordingly, these representations are largely protected from interference. The disruption of hippocampal representations may thus not manifest as graded forgetting but eventually lead to holistic forgetting in an all-or-none fashion (18, 65, 67). Indeed, a previous study that did not observe evidence for suppression-induced forgetting also did not obtain evidence for lingering effects on hippocampal representations (69). Comparisons with this study, however, are difficult, given that it examined the suppression of consolidated rather than freshly formed memories.

The classifier is a powerful tool for harnessing the information contained in activity patterns across multiple voxels to infer the reactivation of category-specific neural representations (70–73). As a downside, it does not provide evidence for the reinstatement of individual exemplars from within a category. Weaker scene evidence during the retrieval of suppressed memories may thus not only reflect reduced reactivation of the respective scene. Instead, it could conceivably also result from greater reactivation of additional, non-specific scene information during the retrieval of baseline memories. However, this interpretation is difficult to reconcile with the observation that, across participants, a greater reduction in scene reactivation for suppressed memories was associated with a stronger decline in vividness.

Nonetheless, in a complementary set of analyses, we further used RSA to track the reinstatement of individual memory representations. Specifically, given that the retrieval of a memory should reinstate its representation, we expected similar activity patterns to emerge for a given memory on the pre- and on the post-test (26, 31, 74, 75). This was the case for the *baseline* condition, where the activity patterns were more similar for the comparison of a memory with itself than with the other memories. By contrast, the reinstatement was not significant for suppressed memories, though we did not obtain evidence that it differed from the baseline effect.

The absence of a significant difference in our study might simply reflect lower power for the more fine-grained, condition-rich analysis of individual activity patterns than for the more generic classifier (56). On the other hand, it might partly reflect that not all participants were equally successful at suppression. Indeed, as for the reactivation scores, we observed a correlation between the neural and the behavioral changes induced by suppression: parahippocampal reinstatement was particularly affected in those people who also experienced the strongest decline in vividness. Together, the two sets of analyses thus support our hypothesis that suppression hinders the neural reinstatement of avoided memories.

Computational modelling suggests that suppression deteriorates memory representations via a targeted inhibition of the respective representation's strongest, i.e., most active, features ((10); see also (27)). These simulations imply that neural representations need to be at least partially reactivated to become liable for disruption (76–78). This account is reminiscent of the non-monotonic plasticity hypothesis (NMPH) that proposes that the strength of a memory is modified according to the degree of its reactivation (33, 79, 80): inactive memories remain unchanged, moderately activated ones get weakened, whereas strongly activated memories will actually be strengthened. On a neurophysiological level, this effect is reflected in synaptic weakening (long-term depression) following moderate postsynaptic depolarization and in synaptic strengthening (long-term potentiation) following a stronger depolarization (79, 81).

Indeed, during initial suppression attempts, unwanted memories often involuntarily start intruding into awareness (13, 82, 83), indicating that they were partly reactivated. Such intrusions then become less frequent over time with repeated suppression attempts. This decrease in intrusions has been associated with a mechanism of reactive inhibitory control that is mediated by an upregulation of the dlPFC and a negative top-down modulation of hippocampal activation (8, 82). It moreover may be reliant on GABAergic activity in the hippocampus (84).

On the one hand, suppression may thus build on an unsupervised learning process as proposed by the NMPH. This process affects memories solely based on the degree of their reactivation and irrespective of any intention to forget (27, 33, 77, 85, 86). On the other hand, this process could be complemented by a supervised top-down process that is mediated by the dlPFC. By disrupting hippocampal retrieval, the top-down process may keep the reactivation of a memory to a moderate level on the plasticity curve and thus render its representation amenable to synaptic weakening.

To conclude, the current study set out to examine the neural consequences of suppression that underlie the phenomenon of suppression-induced forgetting (4, 36, 37). We demonstrated that suppression rendered memories less vivid and, at the same time, also hindered the reactivation of their neural representations. Notably, a weaker reinstatement of the memories was also associated with a greater reduction in vividness. We thus tie, for the first time, the phenomenological changes induced by suppression to their neural basis.

Methods

Participants. Thirty-seven right-handed volunteers participated in this study. They were all drawn from the participant database of the Max Planck Institute for Human Cognitive and Brain Sciences, reported no history of psychiatric or neurological disorder, gave written informed consent as approved by the local research ethics committee, and were reimbursed for their time. Four participants were excluded either due to technical problems (2), non-compliance with the instructions as assessed by a post-experimental questionnaire (1) derived from (87), or drop out (1). We thus included 33 participants in the analysis (age: M = 24.85 y, range = 20-28 y; 17 females, 16 males). We had aimed for a final sample of about 30 participants and thus recruited 37 participants in anticipation of possible exclusions due to non-compliance or excessive movement. This sample size of 30 was based on previous studies studying suppression and our behavioral pilot.

Materials. The stimuli for the experimental procedure were taken from (38). They comprised 60 object-scene pairs: 48 critical pairs and 12 filler pairs. The scenes were negative images depicting traumatic scenes and were originally selected from the International Affective Picture System (88) and online sources. The objects were photographs of familiar, neutral objects taken from (89). Specifically, each object was chosen to resemble an object that was also part of its paired scene (but not essential to the gist of the scene). Throughout the experiment, all images were presented on a grey background. The 48 critical pairs were divided into three sets that were matched on salience of the objects, as well as the emotional valence and arousing nature of the scenes (38). Assignment of the sets to the three conditions was counterbalanced across participants.

The task for training the pattern classifier was based on (27). It included black and white photographs of five different categories: aversive scenes, neutral scenes, morphed scenes, objects, and fruits. The aversive scenes were different items taken from the same databases as the critical items. We created morphed pictures of the aversive scenes with the procedure described by (48). The morphed pictures retain the low-level visual features of the original pictures while ensuring that their content could no longer be recognized. We placed the pictures of the objects and fruits on top of phase-scrambled versions of the scenes and thus ensured that the images had the same size and rectangular shape. All images for the classifier training were normalized with respect to their luminance using the procedure described by (33). The experiment was presented using Psychtoolbox (90, 91).

Procedures.

Experimental design. We tested the impact of suppression on memory reinstatement using an adapted version of the Think/No-Think procedure developed by (38). This procedure entailed four critical phases: an initial study phase, a pre-test, the suppression phase, and a post-test. These were followed by a classifier training task in the scanner and an additional memory task (see supplement). The entire session took around four hours.

During the initial study phase, participants encoded all object-scene associations. First, they saw each object together with its scene, and tried to intentionally remember the associations and, in particular, the scenes in as much detail as possible. Each pair was presented for 6 s followed by a 1 s inter-trial-interval (ITI). Following initial encoding, we presented each object as a cue and asked participants to indicate within 5 s, via button press, whether they could fully recall the associated scene. Once they had pressed the button, they had again 5 s to choose the correct scene out of an array of three different scenes (all of which were drawn from the actual stimulus set). The correct object-scene pair was then presented as feedback. This procedure was repeated up to three times until participants had correctly identified at least 60% of the scenes. To facilitate learning, this phase was split into two parts, each with half of the object-scene associations. Finally, participants were again shown all objects and requested once more to indicate whether they could recall the complete scene without receiving any feedback.

Participants then moved to the MRI scanner. Here, they saw all pairs a final time for 1.5 s each with a 800 ms ITI. The extensive learning regiment and this refresher immediately prior to the critical parts of the experiment ensured that participants had encoded strong associations and were able to vividly recall the scenes although it made it less likely that suppression would induce absolute forgetting rather than gradual fading of the memories (8).

During the pre-test, we presented all 48 reminders on the screen for 3 s each. Participants had to covertly recall the associated scene in as much detail as possible for the duration of the whole trial. They then rated the vividness of their recollection on a scale from 1 (not vivid at all) to 5 (very vivid). We presented no feedback at this stage. The rating was followed by a long ITI of 14 s. With this long ITI, we optimized our ability to detect the activity pattern associated with the recollection of a given scene with little contamination of the activity pattern on the subsequent trial (27). The order of trials was pseudorandomized in a way that at most three objects from the same condition were presented in a row.

The critical suppression phase consisted of five blocks. During a block, each object was presented two times for 3 s. A green frame around an object indicated that participants had to perform the *recall* task. That is, here they had to recall the associated scene as vividly as possible. By contrast, a red frame around an object indicated that participants had to perform the *suppress* task. We here had instructed them to engage a mechanism that we have previously shown to disrupt hippocampal retrieval (7, 9, 10, 13). That is, they were asked to avoid the associated scene from coming to mind while focusing on the object on the screen. If the scene were to intrude into their awareness, they were requested to actively push it out of their mind. Importantly, a third of the objects were not shown during this phase. These items thus served as baseline memories. The ITIs were optimized with optseq (https://surfer.nmr.mgh.harvard.edu/optseq/) and ranged from 2 s to 8.5 s with a mean of 3 s. Participants received extensive training and feedback on this procedure on the filler memories prior to entering the scanner. Immediately following the suppression phase, participants performed the post-test. This phase was identical to the pre-test but with a different pseudorandom presentation order.

Finally, participants also engaged in a classifier training task (modelled on (27)) to obtain a clear neural signal associated with the perception of aversive scenes. We presented pictures of the five categories in separate task blocks. During each block, they saw 10 different pictures of the given category for 900 ms with a 100 ms ITI. Six of these pictures were randomly repeated within each block, thus resulting in 16 trials. Participants had to indicate the occurrence of these repetitions via a button press to ensure that they actually attended to the stimuli. Each category was presented in six blocks (for 30 blocks in total) in a pseudorandom presentation order with no more than two blocks of the same category in a row and with 10 s inter-block-intervals.

Participants also completed a number of questionnaires. In addition to assessing compliance with the instructions, these were designed to assess their strategy use and subjective ratings on recall and suppression success. Further, they filled in Beck's Depression Inventory II, (BDII, (92)), the Thought Control Ability Questionnnaire (TCAQ, (93)) and the State-Trait Anxiety Inventory (STAI, (94)). These data were not analyzed for the current purpose.

fMRI data acquisition. We used a 3T Siemens Prisma MRI Scanner with a 32-channel head coil at the Max Planck Institute for Human Cognitive and Brain Sciences. Structural images were acquired with a T1-weighted MPRAGE protocol (field of view = 256 mm, 1 mm isotropic voxels, 176 sagittal slices with interleaved acquisition, TR = 2300 ms, TE = 2.98 ms, flip angle = 9, phase encoding: anterior-posterior, parallel imaging = GRAPPA, acceleration factor = 2. Functional images were acquired using a whole brain multiband echo-planar imaging (EPI) sequence (field of view = 192 mm, 2 mm isotropic voxels, 72 slices with interleaved acquisition, TR = 2000 ms, TE = 25 ms, flip angle = 90, phase encoding: anterior-posterior, acceleration factor = 3). 369 volumes were acquired in pre- and post-tests, 197 volumes in each suppression block and 395 volumes in the classifier training. The first five volumes of each run were discarded to allow for T1 equilibration effects. Pulse oxy data was collected on participants' left hand. Participants gave their responses via a 5-button box in their right hand.

Analyses.

fMRI data preprocessing. The MRI data were first converted into the Brain Imaging Data Structure (BIDS) format (95). All data preprocessing was performed using the default preprocessing steps of fMRIPrep 1.5.0rc2, based on Nipype 1.2.1. (96): The respective T1 volume was corrected for intensity non-uniformity and skull-stripped, before it was segmented into cerebrospinal fluid (CSF), white matter (WM), and grey matter (GM). It was then spatially normalized to the ICBM 152 Nonlinear Asymmetrical template version 2009c using nonlinear registration.

The functional data were slice-time corrected, motion corrected, and corrected for susceptibility distortions. They were then coregistered to the corresponding T1 image using boundary-based registration with six degrees of freedom. Physiological noise regressors were extracted to allow for component based noise correction. Anatomical CompCor components were calculated within the intersection of the subcortical mask and the union of CSF and WM masks, after their projection to the native space of each functional run. Framewise displacement was also calculated for each functional run. For further details of the pipeline, including the software packages used by fMRIPrep, please refer to the online documentation (https://fmriprep.org/en/20.2.0/). Our univariate analyses were performed in MNI space (following smoothing with a Gaussian kernel of 6 mm FWHM), whereas the multivariate pattern analyses (MVPA) were done on unsmoothed data in native space.

Regions of interest. We manually traced the posterior PhC on the individual T1-weighted structural images, following the anatomical demarcation protocol by (52, 53). Specifically, we defined the PhC as the posterior third of the parahippocampal gyrus (31). We further used the grey matter mask from the fMRIprep pipeline segmented using FSLfast as an ROI.

First-level fMRI analysis. Data were analyzed using SPM12 (www.fil.ion.ucl.ac.uk/spm). We decomposed the variance in the BOLD time series using general linear models (GLM) (97). For the univariate analysis of the suppression phase, we analyzed the data with a GLM including a regressor for the trials of the *recall* condition and a regressor for the trials of the *suppress* condition.

For our multivariate pattern analyses (MVPA), we assessed the individual activity patterns adopting a least-squares-single approach (98). That is, for the pre- and post-test, we estimated separate GLMs for each trial with a regressor for that specific trial and a second regressor for all other trials. For the suppression phase, a given GLM included a regressor coding for all repetitions of the same object and a second regressor for all other trials. For the classifier training task, we estimated separate GLMs for each block with a regressor for that specific block and a second regressor for all other blocks.

All of these regressors coded for the respective 3 s of each trial (or 10 s of each block for classifier training) and were convolved with the canonical hemodynamic response function. In addition, each GLM included six head motion parameters, framewise displacement, the first six aCompCor components and a block regressor as nuisance regressors. We then applied a 128-Hz high-pass filter to the data and the model. For the MVPA analyses, the resulting parameter estimates were transformed into *t*-values via a contrast of the respective individual trial versus all other trials.

Classification analysis. We performed the classifier analysis using the decoding toolbox (47). Specifically, we trained a linear support vector machine for each participant to distinguish activity patterns associated with intact aversive scenes versus their morphed versions. We employed a leave-one-out cross-validation approach that used, on each iteration, eleven of the twelve blocks as training data. This procedure assigns a linear weight to each voxel that reflects its importance in discriminating the two classes, thus creating a weight map. We then used the transformed weight pattern (99) to estimate reactivation as the degree of scene evidence during each trial of the pre-test, post-test and suppression phase. This was done by calculating the dot product of the weight pattern and the respective individual *t*-map.

Representational Similarity Analysis. We examined the reinstatement of unique memory representations using representational similarity analysis (RSA). Specifically, we assessed whether the retrieval of a given scene was associated with a similar neural activity pattern before and after the suppression phase. This analysis used the RSA toolbox (55). It was based on the 48 trials from the pre-test and the post-test. We computed the similarity values using Pearson correlation across all voxels of the respective ROI (54). Specifically, we assessed the similarity of each item with itself (same-item similarity) and the average similarity of the item with all 15 other items from the same condition (different-item similarity). By constraining the different-item similarity to items of the same category, we ensure that any differences with the same-item similarity do not simply reflect general condition differences (i.e. systematic pattern differences for baseline versus suppress items). The similarity estimates were then Fisher-transformed and averaged for each condition within subjects. We determined the magnitude of pattern reinstatement as the difference score between same-item and different-item similarity.

Statistical Analyses. Statistical tests were done with R version 3.6.1 (R Core Team 2019). Repeated measures ANOVAs were conducted with the afex package (Type 3 sums of squares; (100)) and effect sizes are reported as generalized eta squared. Follow-up tests were based on estimated marginal means (lsmeans package, (101)) using pooled variances and degrees of freedom (based on the Welch–Satterthwaite equation). Two-sample *t*-tests (or Wilcoxon signed rank tests in case of skewness) were conducted and effect sizes are reported as Cohen's *d*. The significance level was set to 5%. Only the robust skipped Spearman's correlations were estimated in Matlab (The MathWorks Inc.) using the robust correlation toolbox (102).

Ethics

Human subjects: The Ethics committee of the Medical Faculty, University of Leipzig, Germany, approved the study (protocol number 333/16-ek). Participants provided written informed consent to participate in the study and publish the group results in a scientific journal.

Data and code availability

Because participants did not give consent for their functional MRI data to be released publicly within the General Data Protection Regulation 2016/679 of the EU, these data can only be made available on request to the corresponding author. Aggregated data as well as analyses scripts in R and Matlab will be made available on public repositories.

Funding

A-KM and RGB were funded by a Max Planck Research Group awarded to RGB. The funder had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

A-KM and RGB designed the study. A-KM collected the data and wrote the analysis code. A-KM and RGB analyzed the data and wrote the paper.

Acknowledgements

We thank Philipp C. Paulus for help in implementing the classifier training task, Roxanne Eisenbeis, Johanna Fiebig, Leonie Kanne, and Sarah-Lena Schaefer for assistance in data acquisition, data transcription and scoring, as well as Bernhard Staresina for advice on the manual PhC segmentations.

References

1. Anderson, M. C. & Hulbert, J. C. Active Forgetting: Adaptation of Memory by Prefrontal Control. *Annual Review of Psychology* **72**, annurev–psych–072720–094140 (2021). URL https://www.annualreviews.org/doi/10.1146/annurev-psych-072720-094140.
2. Nørby, S., Lange, M. & Larsen, A. Forgetting to forget: On the duration of voluntary suppression of neutral and emotional memories. *Acta Psychologica* **133**, 73–80 (2010). URL http://linkinghub.elsevier.com/retrieve/pii/S0001691809001498.
3. Fawcett, J. M. & Hulbert, J. C. The Many Faces of Forgetting: Toward a Constructive View of Forgetting in Everyday Life. *Journal of Applied Research in Memory and Cognition* S2211368119301767 (2020). URL https://linkinghub.elsevier.com/retrieve/pii/S2211368119301767.
4. Stramaccia, D. F., Meyer, A.-K., Rischer, K. M., Fawcett, J. M. & Benoit, R. G. Memory suppression and its deficiency in psychological disorders: A focused meta-analysis. *Journal of Experimental Psychology: General* (2020). URL http://doi.apa.org/getdoi.cfm?doi=10.1037/xge0000971.
5. Anderson, M. C. & Hanslmayr, S. Neural mechanisms of motivated forgetting. *Trends in Cognitive Sciences* **18**, 279–292 (2014). URL http://linkinghub.elsevier.com/retrieve/pii/S1364661314000746.
6. Depue, B. E., Curran, T. & Banich, M. T. Prefrontal Regions Orchestrate Suppression of Emotional Memories via a Two-Phase Process. *Science* **317**, 215–219 (2007). URL http://www.sciencemag.org/cgi/doi/10.1126/science.1139560.
7. Benoit, R. & Anderson, M. Opposing Mechanisms Support the Voluntary Forgetting of Unwanted Memories. *Neuron* **76**, 450–460 (2012). URL http://linkinghub.elsevier.com/retrieve/pii/S0896627312007076.
8. Benoit, R. G., Hulbert, J. C., Huddleston, E. & Anderson, M. C. Adaptive Top–Down Suppression of Hippocampal Activity and the Purging of Intrusive Memories from Consciousness. *Journal of Cognitive Neuroscience* **27**, 96–111 (2015). URL http://www.mitpressjournals.org/doi/10.1162/jocn_a_00696.
9. Benoit, R. G., Davies, D. J. & Anderson, M. C. Reducing future fears by suppressing the brain mechanisms underlying episodic simulation. *Proceedings of the National Academy of Sciences* **113**, E8492–E8501 (2016). URL http://www.pnas.org/lookup/doi/10.1073/pnas.1606604114.
10. Gagnepain, P., Henson, R. N. & Anderson, M. C. Suppressing unwanted memories reduces their unconscious influence via targeted cortical inhibition. *Proceedings of the National Academy of Sciences* **111**, E1310–E1319 (2014). URL http://www.pnas.org/lookup/doi/10.1073/pnas.1311468111.
11. Gagnepain, P., Hulbert, J. & Anderson, M. C. Parallel Regulation of Memory and Emotion Supports the Suppression of Intrusive Memories. *The Journal of Neuroscience* **37**, 6423–6441 (2017). URL http://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.2732-16.2017.
12. Paz-Alonso, P. M., Bunge, S. A., Anderson, M. C. & Ghetti, S. Strength of Coupling within a Mnemonic Control Network Differentiates Those Who Can and Cannot Suppress Memory Retrieval. *Journal of Neuroscience* **33**, 5017–5026 (2013). URL http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.3459-12.2013.
13. Mary, A. *et al.* Resilience after trauma: The role of memory suppression. *Science* **367**, eaay8477 (2020). URL https://www.sciencemag.org/lookup/doi/10.1126/science.aay8477.

14. Anderson, M. C., Bunce, J. G. & Barbas, H. Prefrontal–hippocampal pathways underlying inhibitory control over memory. *Neurobiology of Learning and Memory* (2015). URL http://linkinghub.elsevier.com/retrieve/pii/S1074742715002178.
15. Staresina, B. P., Duncan, K. D. & Davachi, L. Perirhinal and Parahippocampal Cortices Differentially Contribute to Later Recollection of Object- and Scene-Related Event Details. *Journal of Neuroscience* **31**, 8739–8747 (2011). URL http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.4978-10.2011.
16. Staresina, B. P., Cooper, E. & Henson, R. N. Reversible Information Flow across the Medial Temporal Lobe: The Hippocampus Links Cortical Modules during Memory Retrieval. *Journal of Neuroscience* **33**, 14184–14192 (2013). URL http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.1987-13.2013.
17. Bohbot, V. D., Allen, J. J. & Nadel, L. Memory Deficits Characterized by Patterns of Lesions to the Hippocampus and Parahippocampal Cortex. *Annals of the New York Academy of Sciences* **911**, 355–368 (2006). URL http://doi.wiley.com/10.1111/j.1749-6632.2000.tb06737.x.
18. Horner, A. J., Bisby, J. A., Bush, D., Lin, W.-J. & Burgess, N. Evidence for holistic episodic recollection via hippocampal pattern completion. *Nature Communications* **6**, 7462 (2015). URL http://www.nature.com/articles/ncomms8462.
19. Tendolkar, I. *et al.* Contributions of the medial temporal lobe to declarative memory retrieval: Manipulating the amount of contextual retrieval. *Learning & Memory* **15**, 611–617 (2008). URL http://www.learnmem.org/cgi/doi/10.1101/lm.916708.
20. Qin, S., van Marle, H. J. F., Hermans, E. J. & Fernandez, G. Subjective Sense of Memory Strength and the Objective Amount of Information Accurately Remembered Are Related to Distinct Neural Correlates at Encoding. *Journal of Neuroscience* **31**, 8920–8927 (2011). URL http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.2587-10.2011.
21. Todd, R. M., Schmitz, T. W., Susskind, J. & Anderson, A. K. Shared Neural Substrates of Emotionally Enhanced Perceptual and Mnemonic Vividness. *Frontiers in Behavioral Neuroscience* **7** (2013). URL http://journal.frontiersin.org/article/10.3389/fnbeh.2013.00040/abstract.
22. Kensinger, E. A., Addis, D. R. & Atapattu, R. K. Amygdala activity at encoding corresponds with memory vividness and with memory for select episodic details. *Neuropsychologia* **49**, 663–673 (2011). URL https://linkinghub.elsevier.com/retrieve/pii/S0028393211000224.
23. Sheldon, S. & Levine, B. Same as it ever was: Vividness modulates the similarities and differences between the neural networks that support retrieving remote and recent autobiographical memories. *NeuroImage* **83**, 880–891 (2013). URL https://linkinghub.elsevier.com/retrieve/pii/S1053811913007283.
24. Xue, G. *et al.* Greater Neural Pattern Similarity Across Repetitions Is Associated with Better Memory. *Science* **330**, 97–101 (2010). URL http://www.sciencemag.org/cgi/doi/10.1126/science.1193125.
25. Ritchey, M., Wing, E. A., LaBar, K. S. & Cabeza, R. Neural Similarity Between Encoding and Retrieval is Related to Memory Via Hippocampal Interactions. *Cerebral Cortex* **23**, 2818–2828 (2013). URL http://www.cercor.oxfordjournals.org/cgi/doi/10.1093/cercor/bhs258.
26. Wing, E. A., Ritchey, M. & Cabeza, R. Reinstatement of Individual Past Events Revealed by the Similarity of Distributed Activation Patterns during Encoding and Retrieval. *Journal of Cognitive Neuroscience* **27**, 679–691 (2015). URL http://www.mitpressjournals.org/doi/abs/10.1162/jocn_a_00740.
27. Poppenk, J. & Norman, K. A. Briefly Cuing Memories Leads to Suppression of Their Neural Representations. *Journal of Neuroscience* **34**, 8010–8020 (2014). URL http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.4584-13.2014.
28. Rissman, J. & Wagner, A. D. Distributed Representations in Memory: Insights from Functional Brain Imaging. *Annual Review of Psychology* **63**, 101–128 (2012). URL http://www.annualreviews.org/doi/10.1146/annurev-psych-120710-100344.
29. King, D. R., de Chastelaine, M., Elward, R. L., Wang, T. H. & Rugg, M. D. Recollection-Related Increases in Functional Connectivity Predict Individual Differences in Memory Accuracy. *Journal of Neuroscience* **35**, 1763–1772 (2015). URL http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.3219-14.2015.
30. Cooper, R. A. & Ritchey, M. Cortico-hippocampal network connections support the multidimensional quality of episodic memory. *eLife* **8**, e45591 (2019). URL https://elifesciences.org/articles/45591.
31. Staresina, B. P., Henson, R. N. A., Kriegeskorte, N. & Alink, A. Episodic Reinstatement in the Medial Temporal Lobe. *Journal of Neuroscience* **32**, 18150–18156 (2012). URL http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.4156-12.2012.
32. Martin, C. B., McLean, D. A., O'Neil, E. B. & Kohler, S. Distinct Familiarity-Based Response Patterns for Faces and Buildings in Perirhinal and Parahippocampal Cortex. *Journal of Neuroscience* **33**, 10915–10923 (2013). URL http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.0126-13.2013.
33. Detre, G. J., Natarajan, A., Gershman, S. J. & Norman, K. A. Moderate levels of activation lead to forgetting in the think/no-think paradigm. *Neuropsychologia* **51**, 2371–2388 (2013). URL http://linkinghub.elsevier.com/retrieve/pii/S0028393213000675.
34. Liu, Y. *et al.* Memory consolidation reconfigures neural pathways involved in the suppression of emotional memories. *Nature Communications* **7**, 13375 (2016). URL http://www.nature.com/doifinder/10.1038/ncomms13375.
35. Wimber, M., Alink, A., Charest, I., Kriegeskorte, N. & Anderson, M. C. Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. *Nature Neuroscience* **18**, 582–589 (2015). URL http://www.nature.com/doifinder/10.1038/nn.3973.
36. Depue, B. E. A neuroanatomical model of prefrontal inhibitory modulation of memory retrieval. *Neuroscience & Biobehavioral Reviews* **36**, 1382–1399 (2012). URL http://linkinghub.elsevier.com/retrieve/pii/S0149763412000383.
37. Anderson, M. C. & Green, C. Suppressing unwanted memories by executive control. *Nature* **410**, 366–369 (2001). URL http://www.nature.com/articles/35066572.
38. Küpper, C. S., Benoit, R. G., Dalgleish, T. & Anderson, M. C. Direct suppression as a mechanism for controlling unpleasant memories in daily life. *Journal of Experimental Psychology: General* **143**, 1443–1449 (2014). URL http://doi.apa.org/getdoi.cfm?doi=10.1037/a0036518.
39. Bergström, Z. M., de Fockert, J. W. & Richardson-Klavehn, A. ERP and behavioural evidence for direct suppression of unwanted memories. *NeuroImage* **48**, 726–737 (2009). URL https://linkinghub.elsevier.com/retrieve/pii/S1053811909006867.
40. Anderson, M. C. Neural Systems Underlying the Suppression of Unwanted Memories. *Science* **303**, 232–235 (2004). URL https://www.sciencemag.org/lookup/doi/10.1126/science.1089504.
41. Depue, B. E., Banich, M. T. & Curran, T. Suppression of Emotional and Nonemotional Content in Memory: Effects of Repetition on Cognitive Control. *Psychological Science* **17**, 441–447 (2006). URL http://pss.sagepub.com/lookup/doi/10.1111/j.1467-9280.2006.01725.x.
42. Karpicke, J. D. & Roediger, H. L. The Critical Importance of Retrieval for Learning. *Science* **319**, 966–968 (2008). URL https://www.sciencemag.org/lookup/doi/10.1126/science.1152408.
43. Karpicke, J. D. & Blunt, J. R. Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping. *Science* **331**, 772–775 (2011). URL https://www.sciencemag.org/lookup/doi/10.1126/science.1199327.
44. Roediger, H. L. & Butler, A. C. The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences* **15**, 20–27 (2011). URL https://linkinghub.elsevier.com/retrieve/pii/S1364661310002081.
45. Linde-Domingo, J., Treder, M. S., Kerrén, C. & Wimber, M. Evidence that neural information flow is reversed between object perception and object reconstruction from memory. *Nature Communications* **10**, 179 (2019). URL http://www.nature.com/articles/s41467-018-08080-2.
46. Dijkstra, N., Ambrogioni, L., Vidaurre, D. & van Gerven, M. Neural dynamics of perceptual inference and its reversal during imagery. *Elife* **9**, e53588 (2020).
47. Hebart, M. N., GÃ¶rgen, K. & Haynes, J.-D. The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics* **8** (2015). URL http://journal.frontiersin.org/article/10.3389/fninf.2014.00088/abstract.
48. Stojanoski, B. & Cusack, R. Time to wave good-bye to phase scrambling: Creating controlled scrambled images using diffeomorphic transformations. *Journal of Vision* **14**, 6–6 (2014). URL http://jov.arvojournals.org/Article.aspx?doi=10.1167/14.12.6.
49. Chang, L. J., Gianaros, P. J., Manuck, S. B., Krishnan, A. & Wager, T. D. A Sensitive and Specific Neural Signature for Picture-Induced Negative Affect. *PLOS Biology* **13**, e1002180 (2015). URL http://dx.plos.org/10.1371/journal.pbio.1002180.
50. Woo, C.-W., Chang, L. J., Lindquist, M. A. & Wager, T. D. Building better biomarkers: brain models in translational neuroimaging. *Nature Neuroscience* **20**, 365–377 (2017). URL http://www.nature.com/articles/nn.4478.
51. Epstein, R., Graham, K. S. & Downing, P. E. Viewpoint-Specific Scene Representations in Human Parahippocampal Cortex. *Neuron* **37**, 865–876 (2003). URL https://linkinghub.elsevier.com/retrieve/pii/S089662730300117X.
52. Insausti, R. *et al.* MR volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. *AJNR. American journal of neuroradiology* **19**, 659–671 (1998).
53. Pruessner, J. C. *et al.* Volumetry of temporopolar, perirhinal, entorhinal and parahippocampal cortex from high-resolution MR images: considering the variability of the collateral sulcus. *Cerebral Cortex (New York, N.Y.: 1991)* **12**, 1342–1353 (2002).
54. Kriegeskorte, N. Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience* (2008). URL http://journal.frontiersin.org/article/10.3389/neuro.06.004.2008/abstract.
55. Nili, H. *et al.* A Toolbox for Representational Similarity Analysis. *PLoS Computational Biology* **10**, e1003553 (2014). URL http://dx.plos.org/10.1371/journal.pcbi.1003553.
56. Nili, H., Walther, A., Alink, A. & Kriegeskorte, N. Inferring exemplar discriminability in brain representations. *PLOS ONE* **15**, e0232551 (2020). URL https://dx.plos.org/10.1371/journal.pone.0232551.
57. Paulus, P. C., Charest, I. & Benoit, R. G. Value shapes the structure of schematic representations in the medial prefrontal cortex. preprint, Neuroscience (2020). URL http://biorxiv.org/lookup/doi/10.1101/2020.08.21.260950.
58. Frankland, P. W., Josselyn, S. A. & Köhler, S. The neurobiological foundation of memory retrieval. *Nature Neuroscience* **22**, 1576–1585 (2019). URL http://www.nature.com/articles/s41593-019-0493-1.
59. Tonegawa, S., Pignatelli, M., Roy, D. S. & Ryan, T. J. Memory engram storage and retrieval. *Current Opinion in Neurobiology* **35**, 101–109 (2015). URL https://linkinghub.elsevier.com/retrieve/pii/S0959438815001270.
60. Liu, X. *et al.* Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature* **484**, 381–385 (2012). URL http://www.nature.com/articles/nature11028.
61. Neunuebel, J. & Knierim, J. CA3 Retrieves Coherent Representations from Degraded Input: Direct Evidence for CA3 Pattern Completion and Dentate Gyrus Pattern Separation. *Neuron* **81**, 416–427 (2014). URL https://linkinghub.elsevier.com/retrieve/pii/S0896627313010854.
62. Knierim, J. J. & Neunuebel, J. P. Tracking the flow of hippocampal computation: Pattern separation, pattern completion, and attractor dynamics. *Neurobiology of Learning and Memory* **129**, 38–49 (2016). URL https://linkinghub.elsevier.com/retrieve/pii/S1074742715001884.
63. Grande, X. *et al.* Holistic Recollection via Pattern Completion Involves Hippocampal Subfield CA3. *The Journal of Neuroscience* **39**, 8100–8111 (2019). URL http://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.0722-19.2019.
64. Parks, C. M. & Yonelinas, A. P. Moving beyond pure signal-detection models: Comment on Wixted (2007). *Psychological Review* **114**, 188–201 (2007). URL http://doi.apa.

org/getdoi.cfm?doi=10.1037/0033-295X.114.1.188.

65. Richter, F. R., Cooper, R. A., Bays, P. M. & Simons, J. S. Distinct neural mechanisms underlie the success, precision, and vividness of episodic memory. *eLife* **5** (2016). URL https://elifesciences.org/articles/18260.

66. Visser, R. M., Lau-Zhu, A., Henson, R. N. & Holmes, E. A. Multiple memory systems, multiple time points: how science can inform treatment to control the expression of unwanted emotional memories. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**, 20170209 (2018). URL http://rstb.royalsocietypublishing.org/lookup/doi/10.1098/rstb.2017.0209.

67. Andermane, N., Joensen, B. H. & Horner, A. J. Forgetting across a hierarchy of episodic representations. *Current Opinion in Neurobiology* **67**, 50–57 (2021). URL https://linkinghub.elsevier.com/retrieve/pii/S0959438820301161.

68. Kumaran, D., Hassabis, D. & McClelland, J. L. What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. *Trends in Cognitive Sciences* **20**, 512–534 (2016). URL http://linkinghub.elsevier.com/retrieve/pii/S1364661316300432.

69. Liu, W., Kohn, N. & Fernández, G. Probing the neural dynamics of mnemonic representations after the initial consolidation. *NeuroImage* **221**, 117213 (2020). URL https://linkinghub.elsevier.com/retrieve/pii/S1053811920306996.

70. Norman, K. A., Polyn, S. M., Detre, G. J. & Haxby, J. V. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences* **10**, 424–430 (2006). URL http://linkinghub.elsevier.com/retrieve/pii/S1364661306001847.

71. Polyn, S. M. Category-Specific Cortical Activity Precedes Retrieval During Memory Search. *Science* **310**, 1963–1966 (2005). URL http://www.sciencemag.org/cgi/doi/10.1126/science.1117645.

72. Haynes, J.-D. A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. *Neuron* **87**, 257–270 (2015). URL http://linkinghub.elsevier.com/retrieve/pii/S0896627315004328.

73. Kuhl, B. A., Rissman, J. & Wagner, A. D. Multi-voxel patterns of visual category representation during episodic encoding are predictive of subsequent memory. *Neuropsychologia* **50**, 458–469 (2012). URL http://linkinghub.elsevier.com/retrieve/pii/S0028393211004088.

74. Danker, J. F., Tompary, A. & Davachi, L. Trial-by-Trial Hippocampal Encoding Activation Predicts the Fidelity of Cortical Reinstatement During Subsequent Retrieval. *Cerebral Cortex* bhw146 (2016). URL http://www.cercor.oxfordjournals.org/lookup/doi/10.1093/cercor/bhw146.

75. Xue, G. The Neural Representations Underlying Human Episodic Memory. *Trends in Cognitive Sciences* (2018). URL http://linkinghub.elsevier.com/retrieve/pii/S1364661318300652.

76. Elsey, J. W. B., Van Ast, V. A. & Kindt, M. Human memory reconsolidation: A guiding framework and critical review of the evidence. *Psychological Bulletin* **144**, 797–848 (2018). URL http://doi.apa.org/getdoi.cfm?doi=10.1037/bul0000152.

77. Sinclair, A. H. & Barense, M. D. Prediction Error and Memory Reactivation: How Incomplete Reminders Drive Reconsolidation. *Trends in Neurosciences* **42**, 727–739 (2019). URL https://linkinghub.elsevier.com/retrieve/pii/S0166223619301511.

78. Lee, J. L., Nader, K. & Schiller, D. An Update on Memory Reconsolidation Updating. *Trends in Cognitive Sciences* **21**, 531–545 (2017). URL https://linkinghub.elsevier.com/retrieve/pii/S1364661317300785.

79. Ritvo, V. J., Turk-Browne, N. B. & Norman, K. A. Nonmonotonic Plasticity: How Memory Retrieval Drives Learning. *Trends in Cognitive Sciences* **23**, 726–742 (2019). URL https://linkinghub.elsevier.com/retrieve/pii/S1364661319301597.

80. Norman, K. A., Newman, E. L. & Detre, G. A neural network model of retrieval-induced forgetting. *Psychological Review* **114**, 887–953 (2007). URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.114.4.887.

81. Bear, M. F. Bidirectional synaptic plasticity: from theory to reality. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **358**, 649–655 (2003). URL https://royalsocietypublishing.org/doi/10.1098/rstb.2002.1255.

82. Levy, B. J. & Anderson, M. C. Purging of Memories from Conscious Awareness Tracked in the Human Brain. *Journal of Neuroscience* **32**, 16785–16794 (2012). URL http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.2640-12.2012.

83. Hellerstedt, R., Johansson, M. & Anderson, M. C. Tracking the intrusion of unwanted memories into awareness with event-related potentials. *Neuropsychologia* **89**, 510–523 (2016). URL http://linkinghub.elsevier.com/retrieve/pii/S0028393216302469.

84. Schmitz, T. W., Correia, M. M., Ferreira, C. S., Prescot, A. P. & Anderson, M. C. Hippocampal GABA enables inhibitory control over unwanted thoughts. *Nature Communications* **8** (2017). URL http://www.nature.com/articles/s41467-017-00956-z.

85. Wang, T. H., Placek, K. & Lewis-Peacock, J. A. More is less: increased processing of unwanted memories facilitates forgetting. *The Journal of Neuroscience* 2033–18 (2019). URL http://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.2033-18.2019.

86. Kim, G., Lewis-Peacock, J. A., Norman, K. A. & Turk-Browne, N. B. Pruning of memories by context-based prediction error. *Proceedings of the National Academy of Sciences* **111**, 8997–9002 (2014). URL http://www.pnas.org/cgi/doi/10.1073/pnas.1319438111.

87. Hertel, P. T. & Calcaterra, G. Intentional forgetting benefits from thought substitution. *Psychonomic Bulletin & Review* **12**, 484–489 (2005). URL http://link.springer.com/10.3758/BF03193792.

88. Lang, P., Bradley, M. & Cuthbert, B. International affective picture system (iaps): affective ratings of pictures and instruction manual. university of florida, gainesville. Tech. Rep., Tech Rep A-8 (2008).

89. Brady, T. F., Konkle, T., Alvarez, G. A. & Oliva, A. Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences* **105**, 14325–14329 (2008). URL http://www.pnas.org/cgi/doi/10.1073/pnas.0803390105.

90. Brainard, D. H. The Psychophysics Toolbox. *Spatial Vision* **10**, 433–436 (1997). URL https://brill.com/view/journals/sv/10/4/article-p433_15.xml.

91. Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision* **10**, 437–442 (1997). URL https://brill.com/view/journals/sv/10/4/article-p437_16.xml.

92. Beck, A. T., Steer, R. A. & Brown, G. Beck depression inventory–ii. *Psychological Assessment* (1996).

93. Luciano, J. V., Algarabel, S., Tomás, J. M. & Martínez, J. L. Development and validation of the thought control ability questionnaire. *Personality and Individual Differences* **38**, 997–1008 (2005). URL https://linkinghub.elsevier.com/retrieve/pii/S0191886904002247.

94. Spielberger, C. D., Gorsuch, R., Lushene, R., Vagg, P. & Jacobs, G. Manual for the state-trait anxiety scale. *Consulting Psychologists* (1983).

95. Gorgolewski, K. J. *et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data* **3**, 160044 (2016). URL http://www.nature.com/articles/sdata201644.

96. Esteban, O. *et al.* FMRIPrep: a robust preprocessing pipeline for functional MRI. preprint, Bioinformatics (2018). URL http://biorxiv.org/lookup/doi/10.1101/306951.

97. Friston, K. J. *et al.* Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping* **2**, 189–210 (1994). URL http://doi.wiley.com/10.1002/hbm.460020402.

98. Mumford, J. A., Turner, B. O., Ashby, F. G. & Poldrack, R. A. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage* **59**, 2636–2643 (2012). URL http://linkinghub.elsevier.com/retrieve/pii/S1053811911010081.

99. Haufe, S. *et al.* On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage* **87**, 96–110 (2014). URL https://linkinghub.elsevier.com/retrieve/pii/S1053811913010914.

100. Singmann, H., Bolker, B., Westfall, J., Aust, F. & Ben-Shachar, M. S. *afex: Analysis of Factorial Experiments* (2020). URL https://CRAN.R-project.org/package=afex. R package version 0.27-2.

101. Lenth, R. *emmeans: Estimated Marginal Means, aka Least-Squares Means* (2020). URL https://CRAN.R-project.org/package=emmeans. R package version 1.4.7.

102. Pernet, C. R., Wilcox, R. & Rousselet, G. A. Robust Correlation Analyses: False Positive and Power Validation Using a New Open Source Matlab Toolbox. *Frontiers in Psychology* **3** (2013). URL http://journal.frontiersin.org/article/10.3389/fpsyg.2012.00606/abstract.