nature COMMUNICATIONS

ARTICLE

Check for updates

https://doi.org/10.1038/s41467-020-18364-1 OPEN

Prior knowledge promotes hippocampal separation but cortical assimilation in the left inferior frontal gyrus

Oded Bein[1], Niv Reggev[2] & Anat Maril[3,4]✉

An adaptive memory system rarely learns information tabula rasa, but rather builds on prior knowledge to facilitate learning. How prior knowledge influences the neural representation of novel associations remains unknown. Here, participants associated pairs of faces in two conditions: a famous, highly familiar face with a novel face or two novel faces while undergoing fMRI. We examine multivoxel activity patterns corresponding to individual faces before and after learning. The activity patterns representing members of famous-novel pairs becomes separated in the hippocampus, that is, more distinct from one another through learning, in striking contrast to paired novel faces that become similar. In the left inferior frontal gyrus, however, prior knowledge leads to integration, and in a specific direction: the representation of the novel face becomes similar to that of the famous face after learning, suggesting assimilation of new into old memories. We propose that hippocampal separation might resolve interference between existing and newly learned information, allowing cortical assimilation. Thus, associative learning with versus without prior knowledge relies on radically different computations.

[1] Department of Psychology, New York University, 6 Washington Pl, New York, NY 10003, USA. [2] Psychology Department, Ben Gurion University of the Negev, 1 Shderot Ben Gurion, Be'er Sheva 8410501, Israel. [3] Department of Psychology, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel. [4] Department of Cognitive Science, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel. ✉email: anat.maril@mail.huji.ac.il

When Madonna releases a new album or starts dating a new person, one cannot avoid the deluge of ads and media posts, and so novel information is added to our knowledge about Madonna. How do we form such new associations? How is learning different when the person is less familiar to us? To be adaptive, learning cannot start de novo each time we form a new association. Rather, we cast our already-existing knowledge to facilitate new learning[1–9]. Indeed, robust behavioral findings demonstrate that prior knowledge facilitates memory of novel associations[2,4,5,9–11]. For example, it is easier to learn that a highly familiar person, such as Madonna, has a new lover than to learn that two unfamiliar people have become lovers[1,6,12,13]. However, while existing memory representations can serve as scaffolding for the assimilation of new information[11,14–19], they can also at the same time produce interference, as multiple existing associations may render the acquisition of new memories more difficult[8,20–22]. In the case of Madonna dating a new person, our previous associated memories—e.g., knowledge about Madonna's previous lovers—might potentially come to mind and interfere with learning the novel association. In the face of this conundrum, the question arises: when forming a new memory, how do we successfully utilize prior knowledge while also protecting against interference?

One possibility is a division of labor, such that prior knowledge biases the cortical memory system toward assimilation of novel information[16], while shifting hippocampal processes to resolve interference[23]. Across rodents and humans, studies demonstrate that during new learning, prior knowledge enhances cortical activation and cortico–cortical functional connectivity, while also modulating hippocampal activation and functional connectivity with the cortex[11,18,19,24–34] (for reviews, see refs. [16,35,36]). Critically, however, univariate activation and functional connectivity studies cannot address how prior knowledge modulates the neural representation of novel associations in these memory systems. As such, in this study, we asked whether the cortical system might support the beneficial effects of prior knowledge on novel learning through assimilation, while the hippocampus defends against potential interference from the same knowledge.

Theoretical frameworks propose that hippocampal processes mitigate interference between novel and existing memories[23]. Indeed, univariate activation in the hippocampus during learning of a novel association reduces forgetting of a previously learned, related association[37]. But how does the hippocampus resolve interference between competing memories? Research in humans and rodents has shown that the hippocampus separates overlapping memories by allocating a distinct activity pattern to each in a process known as pattern separation[38–42] (for a potential role of the hippocampus in pattern completion, namely, the recovery of a complete activation pattern from a partial cue, see Discussion). Pattern separation processes could potentially mitigate proactive interference from existing associations when a novel association is added to a previously known item. If so, we would expect that learning about Madonna's new friend would cause the representations of Madonna and her friend to become more distinct.

While prior knowledge might bias the hippocampus toward pattern separation, it might also lead to assimilation of novel information in cortical regions[17,36]. As noted above, prior knowledge increases univariate activation and cortico–cortical interactions[11,32,33,43,44], lending some support to cortical involvement in prior knowledge influences on learning. However, it remains unclear how assimilation occurs at the level of neural representation. If new information is indeed woven into an existing cortical representation[14–16], we propose that when supported by prior knowledge, the learning of novel associations leads to asymmetric cortical effects. To illustrate, since a cortical representation of Madonna has been stabilized over a lifetime of exposure, it might only change slightly to incorporate the novel association with her friend. The representation of the friend, however, will undergo a disproportionately larger transformation during learning, in order to be assimilated into our existing representation of Madonna.

Of a cohort of cortical regions that are involved in prior knowledge influences on new learning[16], noteworthy are the left inferior frontal gyrus (left IFG), the angular gyrus (AG), and the medial–prefrontal cortex (mPFC). We propose that the left IFG may be a good candidate to mediate the assimilation of novel information. Univariate activation in the left IFG has been consistently shown to promote memory of new information that is related to existing semantic knowledge[33,43,44]. The left IFG is also known to be involved in semantic processing more broadly[45–48]. The AG and mPFC have also been proposed to reinstate schematic knowledge and mediate prior knowledge influences on encoding, and thus might also be cortical sites for assimilation (for reviews, see refs. [16,36]).

To test these ideas, we had human participants learn associations between different pairs of faces. One type of pair comprised a famous and a novel face, whereas the other type of pair comprised two novel faces. Critically, we presented each face alone both before and after associative learning[49–51]. Using a representational similarity analysis[52,53] (RSA), we assess how the multivoxel-pattern representations of individual faces change as a function of learning, and of whether or not they included an association with a famous face, that is, whether or not the pair contained an element of prior knowledge. A final associative memory test is used to determine whether learning-related representational changes contribute to subsequent memory. We show that prior knowledge differentially modulates associative learning in the hippocampus versus the cortex. Namely, prior knowledge leads to representational separation in the hippocampus, but assimilation in the cortex. Together, these findings suggest a candidate mechanism for assimilating new information into existing knowledge structures while reducing memory interference.

Results

During the associative learning task, participants repeatedly observed two types of face pairs: either a famous and a novel face (prior knowledge, PK) or two novel faces (no prior knowledge, n-PK). Before and after learning, each face was presented alone on the screen to enable the capture of its multivoxel-pattern representation (Methods, Fig. 1). In both tasks, participants performed orthogonal gender judgments about the faces. Accuracy during learning and during the pre- and post-learning scans was above 96%, demonstrating that participants complied with task instructions (see Supplementary Fig. 1 and Supplementary Note 1 for further details).

Associative memory test. After the post-learning scan, participants were tested for associative memory for all face pairs in a three-alternative forced choice task. Accuracy rates for both types of face pairs (PK, n-PK) were significantly above chance (33%; PK: $M = 0.46$, $SD = 0.18$, $t_{(18)} = 2.89$, $P = 0.01$, Cohen's $d = 0.66$; n-PK: $M = 0.44$, $SD = 0.14$, $t_{(18)} = 3.24$, $P = 0.005$; Cohen's $d =$ 0.74). Overall accuracy rates did not reliably differ between pair types ($t_{(18)} = 0.41$, $P = 0.69$). Importantly, participants had more high-confident hits ("sure" and "possibly" responses, excluding "maybe" responses) for PK pairs compared to n-PK pairs (PK: $M = 0.35$, $SD = 0.18$, n-PK: $M = 0.23$, $SD = 0.15$, $t_{(18)} =$ 2.88, $P = 0.01$, Cohen's $d = 0.66$; Fig. 1d). Thus, our results are

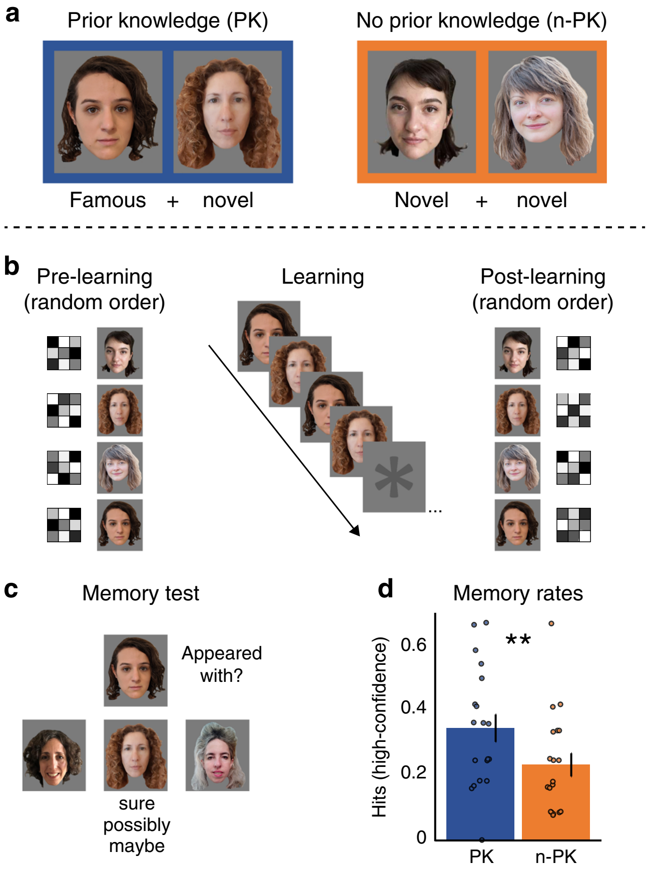

Fig. 1 Design and behavior. a Experimental conditions: participants learned pairs of faces, either a famous and a novel face in the prior-knowledge condition (PK), or two novel faces in the no-prior-knowledge condition (n-PK). **b** While in the scanner, participants viewed the pairs 12 times in 12 cycles; each cycle included all pairs in a random order. Before and after associative learning, participants in the scanner also viewed each face presented alone, in random order. This allowed us to capture the multivoxel activity pattern of each face, for pattern similarity analysis (see Methods). **c** After the post-learning scan, participants performed an associative memory test, in which they indicated which of the three bottom faces appeared with the face on the top and rated their confidence (sure/possibly/maybe). **d** Behavioral results of the associative memory test. High-confidence hits include "sure" and "possibly" responses. $^{**}P = 0.01$, the results of a paired-sample two-tailed t test between PK and n-PK pair type. Data are presented as mean values, error bars reflect +/− SEM. To protect copyrights, all faces in the figures are of novel faces, and we obtained permission to use these photos. Participants in the study saw famous and novel faces, as detailed in Methods.

consistent with previous findings showing that prior knowledge enhances new learning[2,6,12,13,32].

Hippocampus: Prior knowledge leads to representational separation. We tested whether the hippocampal representations of two associated items become more distinct when learning involved prior knowledge. This is in contrast to novel items that do not involve prior knowledge. In this case, a prior study using novel visual fractals has shown that after learning, the hippocampal representations of paired fractals became more similar to each other[50]. Due to the known role of the hippocampus in supporting associative memory (e.g., refs. [54–57]), we predicted that representational changes would be specific to remembered face pairs. To that end, we examined how learning altered the similarity between the multivoxel BOLD activity patterns of items in a pair by computing the change in similarity from before to after learning. We then compared learning-related changes in similarity between members of PK and n-PK pairs, and dependent on whether the association between faces was later remembered (high-confidence hits) or forgotten in the final associative memory test (Fig. 2a).

Similarity differences were submitted to a 2 (prior knowledge: PK, n-PK) by 2 (memory: remembered—high-confidence hits only, forgotten) repeated-measure ANOVA. As shown in Fig. 2, there was a significant interaction between prior knowledge and memory in the left anterior hippocampus ($F_{(1,17)} = 10.12$, $P = 0.0055$, $\eta_p^2 = 0.37$; survives Bonferroni correction for four regions of interest (ROIs), namely, left/right hemisphere by anterior/posterior hippocampus, $P < 0.0125$). This interaction stemmed from face representations in PK pairs becoming more distinct (less similar) from one another after learning, in contrast to representations in n-PK pairs, which became more similar to each other after learning (remembered, PK vs. n-PK: $t_{(17)} = 2.98$, $P = 0.008$, CI: [−0.24–(−0.04)], Cohen's $d = 0.70$). Interestingly, similarity changes only occurred for pairs that were later remembered (significant for PK pairs, remembered vs. forgotten: $t_{(17)} = 2.48$, $P = 0.02$, CI: [−0.14–(−0.01)], Cohen's $d = 0.58$; approaching significance for n-PK pairs, remembered vs. forgotten: $t_{(17)} = 1.92$, $P = 0.07$, CI: [−0.008–0.17], Cohen's $d = 0.45$). Neither significant interactions nor main effects were observed elsewhere in the hippocampus. The data from the other hippocampal ROIs are reported in Supplementary Fig. 5 and Supplementary Note 8.

While this study is suitable to examine the differences from before to after learning[50,51], we did not aim to specifically look at each of the pre-learning or post-learning phases separately, as these similarity values might be tainted by several factors like correlations between the regressors in the general linear model (GLM) used to analyze fMRI data[58,59]. Notably, since the scans before and after learning scans were identical, subtracting similarity values and looking at the differences, as was done in previous studies, solves these issues (see Methods and refs. [49–51]). Nevertheless, to facilitate future research, we examined these values, and we note these results here and in full in Supplementary Note 5. As expected, we found that for remembered pairs, similarity values after learning were qualitatively lower in PK pairs compared to n-PK pairs. We additionally found that before learning, similarity values for remembered PK pairs were qualitatively higher than those of the remembered n-PK pairs, suggesting that prior knowledge might influence the preconditions that render associative learning successful and lead to subsequent memory of associations (see Supplementary Note 5). Future research, using a suitable design that enables a careful examination of the before and after learning values separately, potentially a slow event-related design, or other imaging techniques, could better elucidate these results.

For robustness, and to further ensure that the reported effects in the left anterior hippocampus were specific to pairs that were learned together, rather than a general effect of observing faces with prior knowledge versus without prior knowledge, we compared remembered pairs to shuffled pairs (items from the same PK/n-PK pair type that did not appear together at learning, see Methods). As above, we looked at the difference in similarity from before to after learning. Once again, remembered PK pairs were significantly less similar to each other than were PK-shuffled pairs ($t_{(17)} = 3.05$, $P = 0.007$, CI: [−0.10–(−0.02)], Cohen's $d = 0.72$; PK-shuffled: $M = -0.004$, $SD = 0.04$). Remembered n-PK pairs were more similar than n-PK-shuffled pairs (n-PK: $t_{(17)} = 2.09$, $P = 0.052$, CI: [0.00–0.16], Cohen's $d = 0.49$; n-PK-shuffled: $M = -0.001$, $SD = 0.015$). Thus, in the n-PK condition, our

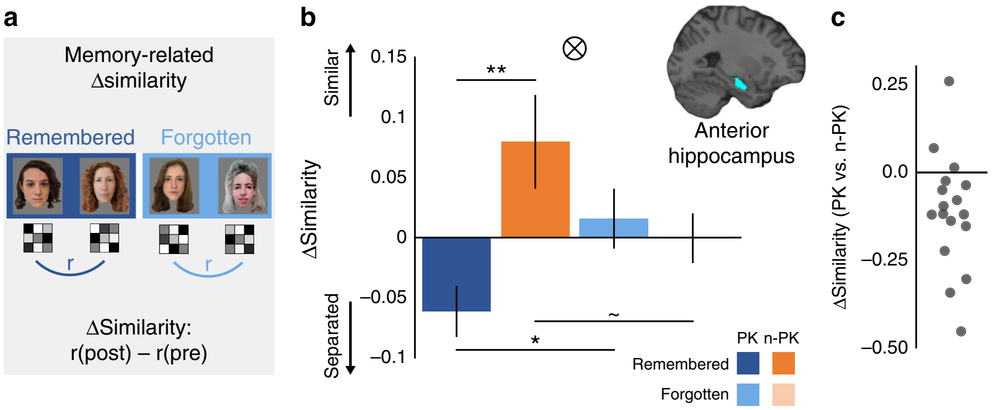

Fig. 2 Representational changes in the hippocampus are modulated by prior knowledge and mediate memory. a Pairs were grouped based on memory in the later associative memory test (remembered pairs include high-confidence hits), and prior knowledge (PK/no prior knowledge, n-PK). The multivoxel activity patterns of items in each pair were correlated before and after learning, and the pre- to post-learning difference in correlation values (Fisher-transformed) was calculated. **b** The results from the left anterior hippocampus. Similarity after learning was lower between members of PK pairs, in contrast to an increase in similarity in n-PK pairs. Similarity differences were specific to remembered pairs. Δ Similarity: difference in similarity values from before to after learning. $N = 18$. Data in the bar graphs are presented as mean values, error bars reflect +/− SEM. ⊗ Interaction of pair type (PK/n-PK) by memory (remembered/forgotten) in a repeated- measure ANOVA: $P = 0.0055$. $^{**}P = 0.008$, $^{*}P = 0.02$, $\sim P = 0.07$, the results of a paired-sample two-tailed t test between pairs of conditions, as depicted by the black lines. **c** Dots reflect individual participants' Δ Similarity reduction due to prior knowledge (PK-remembered pairs minus n-PK-remembered pairs). To protect copyrights, all faces in the figures are of novel faces, and we obtained permission to use these photos. Participants in the study saw famous and novel faces, as detailed in the Methods.

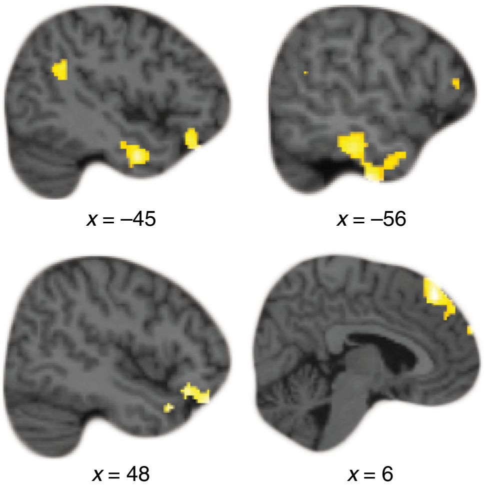

Fig. 3 Functional connectivity with the left anterior hippocampus. Regions demonstrating significantly higher functional connectivity (PPI) with the left anterior hippocampus for prior-knowledge (PK) pairs compared to no-prior-knowledge (n-PK) pairs during associative learning (see also Supplementary Table 1). $N = 19$.

results are consistent with previous research demonstrating increased similarity for novel pairs of items[50]. Critically, our results show the opposite pattern when novel information becomes associated with prior knowledge due to learning—in this case, the items' representations became more separated.

Prior knowledge enhances hippocampal–cortical functional connectivity. If semantic knowledge is represented in the cortex[16,47,60], and hippocampal processes resolve interference between this knowledge and new learning, we reasoned that there should be higher hippocampal–cortical functional connectivity during learning in PK pairs compared to n-PK pairs. Such crosstalk might reflect input of cortical information about the famous faces to the hippocampus, or top–down control signaling the need for interference resolution. To test this, we examined functional connectivity between the left anterior hippocampus, where multivoxel-pattern similarity differences were observed, and the rest of the brain using a psychophysiological interaction analysis[61] (PPI). We compared all PK pairs to n-PK pairs during the associative learning task. Consistent with our prediction, we found a host of cortical regions that demonstrated higher functional connectivity with the anterior hippocampus for PK compared to n-PK pairs, including the left IFG, the AG, and the left middle temporal gyrus (see Fig. 3 and Supplementary Table 1 for detailed results). All of these regions are involved in the processing of famous faces, or more broadly in semantic processing[47,60,62,63]. No significant differences in functional connectivity were observed in the opposite n-PK > PK contrast. These results show that prior knowledge enhances communication between cortical regions and the hippocampus during new associative learning, consistent with our predictions.

We note that no mPFC clusters emerged in this analysis. While not all prior knowledge studies report medial–prefrontal findings (for a review, see ref. [16]), a recent study did find slightly higher hippocampus–mPFC connectivity when participants associated pairs of famous faces and houses, as compared to novel faces and houses[32]. Given the broad interest in hippocampus–mPFC interactions[35,64], and specifically in relation to prior knowledge[16,25,29,36], we examined whether an mPFC cluster would emerge using a more liberal threshold of $P < 0.01$, voxel level. Indeed, a region showing higher connectivity in PK pairs compared to n-PK pairs emerged in the ventral and anterior part of the mPFC ([2, 52, −24], 163 voxels). The opposite contrast of n-PK > PK did not reveal any mPFC cluster at this statistical threshold.

Left IFG: new information is assimilated into cortical knowledge structures. Next, we tested whether prior knowledge led to assimilation in the cortex[17–19,36,65]. Specifically, we asked whether cortical representations showed evidence of asymmetric updating, in which the representation of a novel face after learning became more similar to the original (pre-learning) representation of the famous face it was associated with. This could indicate that the

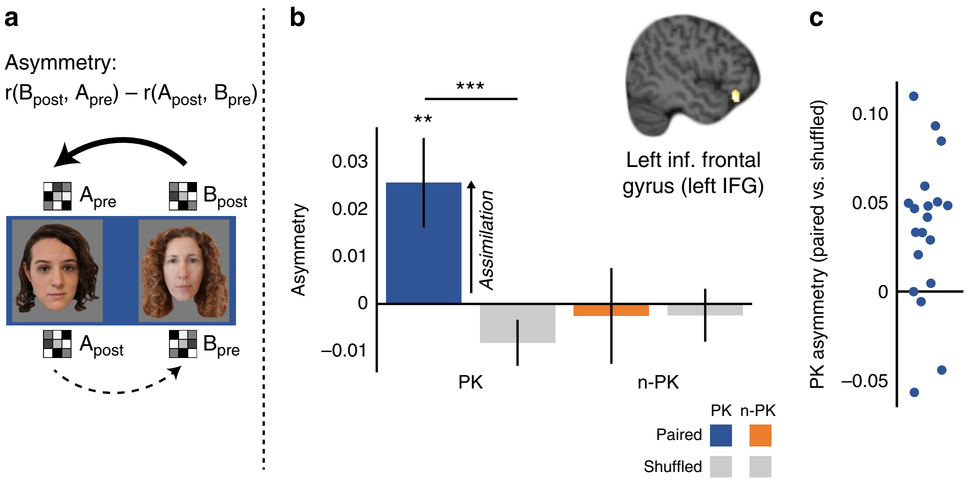

Fig. 4 Asymmetric similarity changes in the left inferior frontal gyrus (left IFG). a Asymmetry in learning reflects the extent to which the representation of the B face after learning became more similar to that of the A face before learning than the representation of the A face that became similar to that of the B face (B_{pre}). In accordance, the multivoxel activity pattern of the B face post learning (B_{post}) was correlated with the pattern of the A face pre-learning (A_{pre}), and the pattern of the A face post learning (A_{post}) was correlated with the pattern of the B- face pre-learning (B_{pre}). Then, the latter similarity value was subtracted from the former as shown above, to reflect the extent to which the B face became more similar to the A face than the A face became to the B face. We interpreted asymmetry as assimilation of the B face into the representation of the A face. **b** Asymmetry was observed in the left IFG in the prior knowledge (PK) pair type for paired faces only, but not in the no-prior-knowledge (n-PK) pair type. $N = 19$. Data in the bar graphs are presented as mean values, error bars reflect +/− SEM. $^{**}P = 0.01$, results of a one-sample two-tailed *t* test against zero. $^{***}P = .002$, results of a paired-sample two-tailed *t* test between PK-paired and PK-shuffled pairs, as depicted by the black lines. **c** Dots reflect individual participants' PK asymmetry difference (PK-paired vs. PK-shuffled). To protect copyrights, all faces in the figures are of novel faces, and we obtained permission to use these photos. Participants in the study saw famous and novel faces, as detailed in Methods.

representation of the novel face was woven into the representation of the famous face (see Introduction and Methods). To the extent that assimilation co-occurs with, or even depends on, hippocampal interference resolution, we reasoned that this should be observed in regions that communicated with the hippocampus during learning. Thus, we chose an ROI that showed higher functional connectivity with the anterior hippocampus for PK than for n-PK pairs (functional connectivity analysis, Fig. 3). We focused on the left IFG due to its roles in mediating the effects of prior knowledge on new learning[33,43,44] (see Introduction).

Although we targeted asymmetric changes in the representational patterns from before to after learning, we first examined whether any changes in representational similarity occurred from before to after learning. To that end, we asked whether there was evidence of learning-related similarity changes from pre- to post learning in the left IFG. For each pair type (PK/n-PK), we compared the change in representational similarity between pairs of faces that appeared together during the learning task to the shuffled-pair baseline (faces from the same PK/n-PK pair type that did not appear together at learning, see Methods). This comparison allowed us to specifically examine changes due to associative learning, controlling for similarity differences due to item familiarity or type of pair[50]. A repeated-measure ANOVA with prior knowledge (PK/n-PK) and pairing (paired/shuffled), revealing a main effect of Pairing ($F_{(1,18)} = 11.44$, $P = 0.003$, $\eta_p{}^2 = 0.39$), and no main effect of prior knowledge or prior knowledge by pairing interaction (*F*'s < 1.2, *p*'s > 0.29). Pairwise comparisons showed a highly significant increase in similarity in PK pairs (PK: paired: $M = 0.03$, $SD = 0.04$; shuffled: $M = -0.01$, $SD = 0.02$; $t_{(18)} = 4.1$, $P = 0.0007$, CI: [0.018–0.058], Cohen's $d = 0.94$). N-PK-paired faces also became qualitatively more similar compared to shuffled baseline (paired: $M = 0.02$, $SD = 0.06$; shuffled: $M = 0.00$, $SD = 0.02$; $t_{(18)} = 1.16$, $P = 0.26$, CI: [−.014–0.049]). Thus, we obtained greater increase in similarity specifically for paired faces, but not for shuffled faces, indicating that there was a change in similarity from before to after learning.

When comparing the representational similarity of items after learning, we only get one measure of the magnitude of similarity between the neural representations of faces in the pairs. It is thus unclear whether similarity increase reflects changes in the representations of both items, such that they both become more similar to each other (symmetric changes), or in one item, which becomes more similar to the other (asymmetric changes). To test our main hypothesis regarding asymmetry in learning, we reasoned that comparing the post-learning pattern of the novel B face to the pre-learning pattern of its paired famous A face would give us a pure measure of how much the representation of the novel B face became more similar to the representation of the famous A face during learning (we performed an identical procedure for novel–novel n-PK pairs). Thus, we compared the representational similarity between the B face after learning (always novel) and the A face (famous or novel) before learning, as well as the representational similarity between the A face after learning and that of the B face before learning. We then subtracted the latter from the former to get an asymmetry measure: if both face representations became equally more similar to one another, there would be no difference between these two values. If, however, the B face after learning became more similar to the A face before learning, while the A face did not become more similar to the B face before learning, then we should observe a positive value, indicating asymmetry[50] (see Methods and Fig. 4). Consistent with our prediction, positive and significant asymmetry was observed only for PK pairs (compared to zero: $t_{(18)} = 2.71$, $P = 0.01$, CI: [0.006–0.046], Cohen's $d = 0.62$, or compared to asymmetry in the shuffled pairs: $t_{(18)} = 3.53$, $P = 0.002$, CI: [0.014–0.054], Cohen's $d = 0.81$, Fig. 4). No asymmetry was found for n-PK pairs (either relative to zero or to asymmetry in shuffled pairs, $t_{(18)}$'s < 0.25, p's > 0.8). Interestingly, we also find some preliminary evidence that asymmetry in the left IFG might be correlated with hippocampal separation (Supplementary Fig. 2 and Supplementary Note 2).

Regarding memory, in the left IFG, the similarity differences from pre- to post learning were not related to subsequent memory in the current study, as no main effect or interaction was obtained in the prior knowledge (PK/n-PK) by memory (remembered/forgotten) ANOVA (all F's < 2.46, p's > 0.13). The asymmetry analysis for remembered and forgotten pairs is reported in Supplementary Note 7.

We further examined whether asymmetry in learning was observed in other cortical regions that demonstrated functional connectivity with the left anterior hippocampus. We chose the AG, since it was recently proposed to bind aspects of schematic knowledge and mediate schema influences on encoding[16,26,66,67]. In the AG, we found no similarity differences from pre- to post learning, with or without respect to subsequent memory (ANOVAs of prior knowledge by paired/shuffled, or of prior knowledge by memory, F's < 2.08, P's > 0.16). For completeness, we have directly targeted asymmetry, which did not differ from zero or from shuffled pairs in either PK or n-PK pairs (all $t_{(18)}$'s < 1.46, P's > 0.16).

As an exploratory analysis, we have examined the other cortical regions that demonstrated functional connectivity with the left anterior hippocampus. The data from these regions are reported in Supplementary Fig. 6 and Supplementary Note 9. Interestingly, even though some of the regions showed an overall increase in similarity from before to after learning, no other region that we examined has shown asymmetry in the direction of representational changes, as was observed in the left IFG.

Ruling out alternative explanations. We emphasize that our results are unlikely to reflect some general response to famous faces during the pre- or post-learning scans. First, we report the difference between the pre- and post scans, and pattern similarity was computed with famous faces in both. Second, critically, our control comparisons in all of the similarity analyses were within condition: shuffled pairs in the PK condition only included pairs in the PK condition, and likewise for the n-PK condition. Nevertheless, the results were specific to paired faces. Another possibility is that the novel faces that were paired with famous faces in PK pairs carry with them some unique status due to becoming associated with famous faces, or that merely being paired with famous faces led to some overall change in the representation of these faces. However, again, any such status that is unrelated to associative learning would have been observed in the shuffled pairs as well. Regarding the hippocampal results, pairs were separated to remembered versus forgotten associations within each pair type, and our results were specific to the remembered faces, alleviating the above concerns for the memory analysis as well.

We further controlled for potential differences in univariate activation during the pre- and post-learning scans[68–70]. As mentioned above, differences in univariate responses between PK and n-PK pairs, if they arose, should have influenced paired and shuffled items alike. Nevertheless, we performed an additional control analysis, in which we included the univariate activity together with the factors of prior knowledge (PK/n-PK) and pairing (and memory, where relevant) in a multiple linear regression. All of the analyses reported here hold when controlling for univariate activation (Supplementary Note 3). Thus, our representational similarity effects are unlikely explained by univariate activity.

Discussion

Here, we asked how new associations are represented in the human brain, and how individual associations are different when adding new to old memories versus learning thoroughly novel associations. Decades of behavioral research have shown the power of prior knowledge to facilitate new learning, such that adding a novel piece of information to existing knowledge is typically easier than thoroughly new learning[1–6,12,71]. In recent years, research has shown that prior knowledge increases cortical activation and functional connectivity, and modulates hippocampal activation and hippocampal–cortical functional connectivity (for reviews, see refs. [15,16,35,36]). However, as previous studies only looked at the strength of brain activation and lesions, they could not address the neural representations of new associations. Thus, the question of whether prior knowledge promotes separation versus assimilation in different neural systems remained unexplored to date.

Theoretically, prior knowledge could facilitate new learning through multiple processes. Prior knowledge is thought to serve as scaffolding for learning by providing an existing cortical representation into which new information can be assimilated. However, existing associations can also interfere with new learning, causing difficulty to associate an item with novel information (i.e., fan effect[8,20–22]). Thus, prior knowledge may promote mechanisms aimed at mitigating interference, such as hippocampal pattern separation. To address this possibility, we examined how prior knowledge altered the neural representations of newly learned associations between either a famous and a novel face or two novel faces.

We found that prior knowledge led to greater separation of the underlying neural representations in the hippocampus. Multivoxel activity patterns of members of famous–novel pairs became less similar to each other after associative learning, whereas representations of novel–novel face pairs became more similar to each other after associative learning. Importantly, these learning-dependent changes in similarity were specific to face pairs that participants later remembered, and did not occur for forgotten pairs. We note that the result in the novel–novel face pairs is consistent with a previous study that found increased similarity in the hippocampus after learning of pairs of novel visual fractals[50]. Critically, this previous study did not address prior knowledge, nor were the previous findings related to subsequent memory in that study. In contrast to the hippocampus, prior knowledge led to cortical assimilation, expressed in asymmetric representational changes in the left IFG. Specifically, we found that the neural representations of novel faces following learning became more similar to the representations of their associated famous faces before learning. Together, these findings show a flexible and directional creation of associations in the human brain that is specific to memory systems and is highly modulated by the state of prior knowledge.

We found that associative learning processes in the hippocampus were highly dependent on prior knowledge. Consistent with previous findings, the representations of novel pair members became more similar to one another after associative learning (ref. [50]). In contrast, we found that famous–novel pair members became more separated after learning. This increased separation supports the idea of interference resolution, in line with previous research on pattern separation (e.g., refs. [40–42,72]). Importantly, separation mediated successful associative memory specifically when prior knowledge was involved, and there was a need to overcome interference from previous associations. Future research can investigate precisely how previous associations interfere with new learning, and how separation processes might facilitate resolution of this interference.

The exact manner by which prior knowledge drives the hippocampus toward separation versus similarity, or integration, is currently unknown. One possibility is that top–down control signals in the prefrontal cortex shift hippocampal computations toward separation. A wealth of research suggests that prefrontal–hippocampal interactions mediate cognitive control processes that select representations for encoding or retrieval from memory (see, e.g., refs. [73,74] for reviews). These control processes might promote interference resolution by biasing hippocampal representations[75–78]. Supporting this possibility, we saw that prior knowledge leads to greater hippocampal–prefrontal interactions during learning and greater subsequent separation in hippocampal representations thereafter.

Another possibility is that differential neuromodulatory input to the anterior hippocampus[79,80] biases the hippocampus toward separation versus similarity[81–83]. Kafkas and Montaldi[79] recently proposed that different types of novelty, such as absolute or contextual, are both detected in the anterior hippocampus, but with different neurotransmitters mediating each type. In our study, novel–novel pairs may evoke an absolute novelty signal, because neither image has ever been seen before. In contrast, new associations involving prior knowledge might promote a contextual novelty signal, because the novel face is novel in the context of the highly familiar face. It has been proposed that absolute novelty enhances acetylcholine input to the hippocampus, while contextual novelty involves the release of dopamine and norepinephrine[79,84–86]. Different neurotransmitters might further lead to separation versus similarity in the hippocampus (e.g., refs. [81,87–89]), as was observed here.

While we found neural signatures of both separation and integration in the anterior hippocampus, it has recently been proposed that these different computations may be localized to the posterior versus anterior hippocampus, respectively[90,91]. Following a previous recent study showing separation in the anterior hippocampus[92], we interpret our findings within a framework embedding the hippocampus in a larger functional network. Anatomically, the anterior hippocampus receives preferential input from the perirhinal cortex (via the entorhinal cortex[93,94]). The perirhinal cortex supports conceptual semantic knowledge and is involved in the processing of items and their features[95–100], potentially as a part of a larger antero-temporal network[101]. In this context, it is not surprising that in our study, which involved associating items and incorporated semantic knowledge, we found representational changes in the anterior hippocampus. Our findings thus suggest that the anterior hippocampus might mediate both separation and integration, dependent on internal knowledge.

The type of prior knowledge involved in new learning could be critical in shifting the hippocampus toward a separation versus an integration mode. Using an associative inference paradigm, a previous study showed that after an A–B pair was learned, learning of an overlapping A–C pair resulted in greater pattern similarity between the B and the C items in the anterior hippocampus[51] (letters represent different items). Interestingly, this was only true if A–B pairs were repeated multiple times prior to learning the A–C pairs. If the learning of the A–C and A–B pairs was interleaved, the anterior hippocampus showed separation. Although they used merely visual associations, one can conceptualize the A–B association as some prior knowledge to which a novel A–C association is added. As one could also think about the famous faces as have been learned over multiple repetitions prior to our study, our separation finding might seem to diverge from the previous results, showing similarity when A–B pairs were repeatedly learned prior to the A–C learning. This might point toward factors that bias hippocampal representations. For example, the A–B pairs form a single association learned over only a few repetitions, whereas the knowledge about the famous faces that we used in this study is highly learned and involves a rich network of strong associations. Thus, one or a few weaker prior associations as in the case of the A–B pairs might not interfere with new learning and result in similarity[51], while multiple strong associations require interference resolution and necessitate separation. However, it may also be the case that our presentation was more similar to the interleaved learning in Schlichting et al.[51], since we reactivated the prior knowledge by presenting the famous face, along with presenting the novel face. Future research could examine how the type of knowledge as well as the learning protocol modulate hippocampal computations.

Another potentially interesting factor is consolidation, or time. Here, we used knowledge that was acquired long before our study and was well consolidated, while in the associative inference study, the A–B associations were learned immediately before the A–C associations. Indeed, a 24-h delay between learning the old A–B and the novel A–C association reduces associative inference[102]. Previous studies have further shown that prior knowledge established immediately before new learning does not enhance associative memory to the same extent as long-held, pre-existing knowledge[103]. Future research should elucidate how the time difference between the initial acquisition of knowledge and the addition of novel associations modulates associative learning in the hippocampus.

In light of the finding discussed above[51,102], it is interesting to consider that prior knowledge might promote pattern completion to retrieve previous related memories, but then separation processes encode the novel association[84,104,105]. Pattern completion refers to the retrieval of a previously encoded activity pattern from a partial cue[106–109]. Prior knowledge may facilitate hippocampal pattern completion to mediate the retrieval of pre-existing-related association. Indeed, univariate activation is typically observed for famous faces compared to novel faces in the hippocampus (for reviews see refs. [60,63]). Nonetheless, this might not necessarily indicate that pattern completion occurs in the hippocampus. It is plausible that hippocampal activation drives pattern completion in other cortical regions where information might be stored[110–113]. It is further important to distinguish between pattern completion of a previous association and encoding of novel associations. Here, we have shown that separation might underlie the addition of a novel association to prior knowledge. One interesting possibility is that pattern completion may mediate the retrieval of previous associations and drive the hippocampal representations of the famous face, and then pattern separation may come into play and mediate the encoding of the novel association—to distinguish the novel association from the previous memories. Otherwise, had the hippocampus remained in pattern completion “state,” interference with previous memories could have happened. Previous theoretical models and empirical work has shown that the switch from a retrieval “state” that is potentially mediated by pattern completion, to an encoding “state” that might require pattern separation, can be rapid in the hippocampus, and might be driven by the relative novelty elicited by the novel face in comparison to the famous face[83,104,114–116].

Researchers have investigated the processes as well as the timeline that characterizes cortical learning for decades (e.g., refs. [19,23,47,117–122]). Here, we investigated the content of individual associations in the cortex[92], and whether the existence of knowledge changes the direction of association, in a specific and predictable way. Theoretical accounts suggest that through learning, new information becomes assimilated into cortical knowledge structures[14,15,17,36]. We hypothesized that assimilation should manifest as asymmetric learning, in which the cortical representation of this new information becomes similar to prior knowledge after learning, while the prior knowledge changes to a lesser extent. Critically, while previous empirical findings of cortical activation and functional connectivity lend some support to the cortical assimilation idea[11,32,33,43,44], they could not address whether and how assimilation occurs at the level of the neural representation. Consistent with our predictions, we found that in the left IFG, representations of the novel faces after learning became more similar to representations of the famous faces before learning (see Methods and Fig. 4), more so than representations of the famous faces after learning became to representations of the novel faces before learning. Importantly, we did not observe such asymmetry in learning when both faces were novel (see also ref. [123] for a related behavioral finding, suggesting that prior knowledge facilitates asymmetrical associations). We propose that asymmetry in learning reflects the assimilation of new information into existing knowledge. This semantic knowledge is acquired across multiple encounters, so existing representations are not modified as strongly when learning additional novel information. Meanwhile, representations of novel information undergo large transformations as they become woven into existing schemas.

Cortical assimilation might be influenced by whether novel information is consistent, inconsistent, or arbitrary with respect to prior knowledge. Here, we show assimilation based on arbitrary associations. Thus, our results are consistent with theories implicating assimilation as a general mechanism for knowledge-supported learning[1,15,16,65,124]. Note, however, that others have proposed that assimilation specifically mediates learning of information that is consistent with prior knowledge[17,18,36]. Similar to our view, these latter frameworks rely on the assumption that the hippocampus is required to prevent interference between new information and cortical knowledge[23]. Thus, when novel information is consistent and elicits less conflict with cortical knowledge structures, cortical assimilation can occur, and hippocampal involvement is reduced[17,36]. In this study, we show that cortical assimilation can occur in parallel with hippocampal involvement. We thus propose that interference resolution occurs either because the novel information elicits less interference to begin with, as in the case of schema-consistent information, or because the hippocampus contributes to the resolution of interference, as in our study. How neural systems may cooperate to determine the neural representation of new memories, and how these processes are shaped by consistency with prior knowledge, are fascinating questions for future research[35].

While in the current study, we employed a memory test in which participants explicitly indicated the paired faces from the learning phase, other measures of memory might be more suitable for unveiling the benefits of neural assimilation. We note that using an explicit memory test, assimilation in the left IFG did not differ between subsequently remembered and forgotten famous–novel pairs (see Results and Supplementary Note 7). This is a null result and thus should be interpreted with caution. However, an interesting possibility is that other, potentially more implicit measures might better uncover the advantages of assimilation. For example, in associative priming paradigms, a presented item leads to a facilitation in the response to an ensuing item, as a result of these items being associated in memory[125]. If the neural representation of the novel face is assimilated into that of the famous face, it might be that upon encountering the famous face, the novel face is spontaneously reactivated, due to the similarity in their neural representation. That could potentially facilitate the response to that novel face. Indeed, we have recently shown that prior knowledge enhances associative priming, using a similar learning paradigm to the one employed here[12]. The explicit memory judgments that participants made in the current study likely relied on hippocampal representations and benefited from retrieval strategies that might overshadow the potential spontaneous reactivation that may have happened in the left IFG. This is consistent with the view that implicit and explicit forms of memory might rely on different mechanisms, and are potentially mediated by different neural systems[57,126,127]. The neural mechanisms by which prior knowledge might influence different forms of learning and memory are an exciting avenue for future research[12].

While we interpret our asymmetry finding as assimilation, it is also possible that after learning, the second face in the pair (B face) brings to mind the first face (A face) more so than vice versa. A previous study found that after sequential learning, the representation of the first item in a pair following learning became more similar to the representation of the second item in the pair before learning[50]. This was interpreted as the first item bringing to mind, or predicting, the second item due to their temporal contingency. In contrast, we found that the second face after learning became similar to the first face before learning. However, applying Schapiro et al.'s[50] interpretation to our findings does raise the possibility that the asymmetry effect we observed reflects the B face bringing to mind the A face. We find this interpretation less likely in our case, because asymmetry was observed in the opposite direction from the temporal order of learning. Moreover, for asymmetry to arise, the B face should elicit the A face more so than the A face elicits the B face. While intriguing, it seems less reasonable that the B face, both novel and temporally second, would make a stronger cue than the A face, which is a famous face and was temporally first in the pair during learning. We thus find the assimilation interpretation more plausible, but acknowledge that the alternative retrieval interpretation should be tested.

Another possibility is that the novel B face brings to mind the famous A face because it is well-known and namable, while the novel B face is not nameable and harder to bring to mind upon seeing the A face. For example, it may be that during the post-learning scans, when seeing the B face that was paired with Madonna, the participants thought "Madonna," but in contrast, the novel and unnamable B face was not recalled upon seeing Madonna. In our view, this possibility is less likely. First, if the participants were bringing to mind Madonna when they saw the B face that was paired with Madonna, that should have reflected on the overall similarity changes from pre- to post learning. However, differences in similarity changes from pre- to post learning between PK and n-PK were far from being significant in the left IFG. Second, if indeed participants recalled Madonna's name upon seeing the B face, we would expect to see asymmetry in other regions as well. For example, in the left AG, or the hippocampus–brain regions known to be involved in recollection (e.g., refs. [128–130]). However, these regions did not exhibit asymmetry, nor did the other cortical regions we examined (see Supplementary Fig. 6 and Supplementary Note 9). Thus, we believe that this interpretation is less likely than the assimilation interpretation. Further, even if indeed the B face brings to mind the famous A face, it might be that this process facilitates assimilation, as it renders the B-face's representation similar to the famous A face. Future research could better elucidate the processes by which assimilation arises.

To conclude, we asked what does it mean to say, "an association was created"? Importantly, an adaptive learning system does not start any learning experience tabula rasa, but rather it utilizes what it already knows about the world. However, reliance on prior knowledge is a double-edged sword, as existing memories can not only enhance but also impair new learning[1,2,5,6,8,20–22,71,131]. While many important questions remain open, our findings suggest a novel putative mechanism for learning as it typically occurs in our everyday lives: we usually add new information to what we already know. In this case, we propose that new information is assimilated into our prior knowledge in the cortex, while hippocampal pattern separation mitigates interference between new and old memories.

Thus, this study clearly demonstrates that associative learning is flexible and directional, specific to memory systems, and highly dependent on prior knowledge.

Methods

Participants. Nineteen right-handed native Hebrew speakers participated in the study (nine women: mean age: 26.94 years, range: 22–31 years). Five additional participants were excluded from the analysis: two due to excessive movement (more than 3 mm across all pre-learning, post-learning, and associative learning scans), two due to insufficient knowledge about the famous faces, as defined by familiarity with fewer than two-thirds of the faces in a post-experiment questionnaire, and one due to poor compliance with the task instructions, leading to lower-than-chance performance in the final memory test. All participants had normal or corrected-to-normal vision, and no color blindness. They were screened to ensure that they had no neurological conditions or any other contraindications for MRI. Participants were paid 280 shekels (equivalent to ~$77) for the study. They were recruited from the Hebrew University of Jerusalem community and provided written informed consent prior to participating in the experiment, in a manner approved by the Tel Aviv Sorasky Medical Center Ethics Committee and The Hebrew University institutional review board.

Materials. Twelve faces of famous women and 36 faces of novel women were used in this study. We used famous faces as our manipulation of prior knowledge because famous faces were shown to elicit a rich representation of previous knowledge (for reviews, see refs. [60,63]) and were used in previous studies examining prior knowledge influences on new associative learning[12,32]. We further followed a previous multivariate fMRI study using female faces to capture the representation of knowledge about faces[132]. The famous faces depicted well-known international and Israeli individuals from a range of fields, including politicians, musicians, actors, and fashion models. An extensive pilot study verified that these faces were indeed familiar to the Hebrew University population, and that the participants could identify them by name and provide details about them. The novel faces were obtained from the Web and included foreign corporate executives, actors, and models that were unfamiliar to our Israeli participants, while controlling for factors such as attractiveness and image quality. Of the 36 novel faces, 12 were selected to match the famous faces with respect to age. For convenience, we refer to the 12 famous faces and the 12 matched novel faces as *A faces*. The remaining 24 novel faces are referred to as *B faces*. The experiment included pairs of face that were presented together, each pair comprising one A face (either famous or novel, see below) and one B face (always novel). The pairing of the each of the 24 novel B faces with each of the 24 A faces (12 famous and 12 novel) was random for each participant. Within each participant, the pairings were fixed throughout the experiment (i.e., for each participant, each A face appeared with one B face, and the same pair appeared in all the repetitions, see below). To enable the associative learning task (see the section Procedure below), we added 6 female faces (3 famous) and 12 novel male faces that comprised mixed-gender pairs.

All the stimuli were color photos of faces presented in the center of a gray rectangle that was 290 pixels (width) by 320 pixels (height). The screen resolution was set to 1024 by 768. To further control for potential visual differences between the pictures, we equated pixelwise similarity (the correlation across the pixel values between the stimuli[133,134]). Since we used color images, we correlated the RGB values of the stimuli with one another, each color layer separately, and averaged the correlation coefficient of each pair of stimuli across the three layers. We also computed pixelwise similarity using grayscale versions of the images, in accordance with previous studies. Overall, the correlation values did not differ between the two measures, indicating the viability of pixelwise color similarity as a measure of pixelwise similarity[12].

A few types of pixelwise similarity were equated across the stimuli. First, we ascertained that the famous faces were equally distinct from one another and from their matched novel A faces. We computed the pixelwise similarity between each of the famous faces and the remaining famous faces and followed the same process for the novel A faces. We obtained similar means and standard deviations for pixelwise similarity across the conditions. Next, we verified that on average, the visual similarity between the famous A- and B faces was equal to that of the novel A- and B faces. To that end, we computed the pixelwise similarity of all the famous faces with all the novel B faces and the similarity of the novel A faces with all the novel B faces. Once again, we obtained similar means and standard deviations for pixelwise similarity across the conditions.

Procedure. The experiment started with pre-learning scans, enabling us to capture the multivariate activity pattern of each face alone prior to learning. This was followed by an associative learning session and post-learning scans, to capture the stimuli patterns after learning. Then, a surprise associative memory test was administered. Critically, acquiring the post-learning scan before testing memory allowed us to measure post-learning representations without interference from probing of memory. The test was followed by an irrelevant task that was not analyzed. All phases were performed in the scanner, and each phase was preceded by detailed instructions and a few practice trials. The presentation of the stimuli was controlled by Presentation® software (Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). Upon completion of all tasks, the participants left the scanner and completed a knowledge questionnaire about the faces that appeared in the experiment and a short debriefing session.

Pre-/post-learning scans. The pre- and post-learning scans were identical[49–51]. All faces that appeared in the learning phase appeared in these scans. In each trial, a face appeared alone at the center of the screen for 1 s. Trials were jittered with 0.5–7.5 s of a fixation-cross baseline, with an interval of 0.5 s, using optseq2 (https://surfer.nmr.mgh.harvard.edu/optseq/[135]). Participants were asked to indicate by pressing a button whether the person appearing on the screen was male or female.

Each phase (pre and post) was divided into two scans; in each scan, each face appeared three times. The order of stimulus presentation was pseudorandomized to maintain low autocorrelations between regressors, and to ensure that two faces that appeared as a pair in the associative learning task appeared with a minimal gap of two stimuli in the pre/post scans, to prevent additional learning during these scans. To create the pseudorandomized order, placeholders of stimuli were fixed, e.g., a certain face appeared in the triad of placeholders at locations 20, 150, and 180. Placeholder triads were then paired such that two faces that would be associated later would each appear in one placeholder triad of the pair (e.g., placeholders 20, 150, and 180 were paired with placeholders 40, 105, and 240). We had two of these fixed orders (determined by simulations to ensure low correlations), one for each scan, and the order of the scans was counterbalanced across participants. To counterbalance the conditions, the pairs of placeholders within each scan were divided into two groups of 12 placeholder pairs. The allocation of famous and novel A faces (and the corresponding B faces) to placeholder groups was counterbalanced across participants. Within each placeholder group, the allocation of placeholders to either A or B faces rotated across participants. The allocation of the stimuli to placeholders was randomized within each condition (famous/novel and A/B face) for each participant. Critically, the pre- and post scans were identical within each participant (we also repeated the same order of scans), and the pattern similarity before learning was subtracted from the pattern similarity after the learning. Thus, differences in pattern similarity cannot be attributed to differences in the correlations between regressors[50]. All faces appeared once before a new cycle of repetition began. Although we analyzed only female faces that appeared in the same-gender pairs during learning, the stimuli from the mixed-gender pairs were included in the pre-/post scans as well, to equate familiarity of the stimuli during the associative learning task, and to enable the male/female gender task during the pre-/post scans. The placeholders of these additional faces were fixed across participants to distribute the males throughout the task, but the allocation of the faces to placeholders was randomized for each participant.

Associative learning task. Participants were presented with pairs of faces that were composed of either a famous and a novel face (PK) or two novel faces (n-PK). In each trial, the faces were presented at the center of the screen. The faces were presented sequentially and not simultaneously to prevent participants from fixating on one face more than the other. Each trial included a double repetition of the pair (A–B–A–B), with each face appearing on the screen for 500 ms (this presentation time ensured recognition of the famous faces[136–138]) and an interstimulus interval (ISI) of 100 ms. A fixation cross appeared at the end of each trial for 600 ms. As before, trials were further jittered with 0.5–7.5-s fixation-cross baseline, with an interval of 0.5 s[135]. The participants had to indicate by pressing a button whether the two faces were two females or a male and a female, and were instructed to respond as quickly and as accurately as they could.

Throughout the learning phase, each pair was repeated 12 times (12 repetitions of A–B–A–B trials). This task was divided into four scans, each of which included three presentations of all pairs. Each cycle of repetition included all 24 experimental pairs and the additional 6 different-gender pairs, which also repeated 12 times throughout the experiment. In these filler trials, the male always appeared second (as a B face). To allow enough males for the pre-/post-learning scans, each female was paired with two males, and these appeared alternatively (such that each male appeared six times in total during the associative learning task). In each cycle, the order of stimuli was pseudorandomized in a similar manner to the pre-/post-learning scans: placeholders were fixed and divided into two groups (for the two conditions, PK and n-PK). The allocation of the groups was counterbalanced across participants. Within each group, the allocation of a specific pair to the placeholders was randomized for each participant. We had four such fixed orders (determined again by simulations to ensure low correlations between regressors), one for each scan, and the order of the scans was randomized for each participant. The placeholders of the female–male pairs were fixed to allow distribution of these trials throughout the task. All pairs appeared once before a new cycle of repetition began.

Associative memory test. Upon completing the associative learning phase, participants performed the post-learning scans. Then, a surprise memory test was given. In each trial, participants were presented with an A face that appeared at the top of the screen (either famous or novel) and three B faces that appeared at the bottom of the screen, one of which had been paired with the A face during the learning session. All three faces were intralist within the pair type (i.e., if an A face was famous, the two distractors were B faces that appeared with other famous faces). The allocation of the distractors was pseudorandomized such that a triad of the same B faces could not appear twice throughout the test. One-third of the B faces appeared as targets in their first presentation, one-third appeared as targets in their second presentation, and one-third appeared as targets in their third presentation. Within each condition (PK/n-PK), the location of the target was equally divided between the three possible locations, and each B face appeared once as a target and twice as a distractor.

In each trial, participants were asked to choose the B face that had appeared with the A face during the learning session. After a face was chosen, the other two faces disappeared, and participants were asked to make a three-level confidence judgment (sure, probably, or maybe, corresponding to high, medium, or low confidence that the faces appeared together, respectively). Both stages of each trial were self-paced, but each stage was limited to 10 s. A 500-ms fixation cross appeared between trials. The order of the trials was randomized such that no more than two trials of either the PK or the n-PK condition appeared consecutively. Within each condition, trials were randomized, and B faces were allocated as distractors such that no face would appear in two consecutive trials, either as target or as distractor.

Knowledge questionnaire. After scanning, participants completed a knowledge questionnaire. All faces appeared one after the other, and subjects had to say whether they knew their names or were familiar with them before the experiment. They were additionally asked to rate how many facts they knew about the person whose face was presented. Since we piloted the famous faces to ensure that people had knowledge about them, this questionnaire was only meant to crudely assess the knowledge of the specific participant. Thus, we excluded subjects that did not recognize (i.e., could not provide the person's name or reported that the person was not familiar to them) over a third of the famous people in the study (two participants). We further excluded from all analyses specific famous faces that were not familiar to a particular participant (six participants each had one face excluded, a different famous face across these participants). Then, participants were debriefed and asked whether they had suspected that there would be a memory test or tried to memorize the pairs during the learning phase.

fMRI parameters and preprocessing. Participants were scanned in a 3T Siemens Prisma scanner. The experiment included an MPRAGE anatomical scan (1X1X1mm resolution), a fieldmap scan, and 12 whole-brain T2*-weighted EPI scans (TR = 2000 ms, 200 × 180-mm FOV, 64 × 58 matrix, TE = 28, flip angle = 77, and phase-encoding direction: anterior–posterior). In each volume, 39 slices were acquired tilted minus 20° of the AC–PC, 3.125 × 3.125 × 3.1-mm (width × length × thickness) voxel size, no gap, and in a top–down (bottom-up for two participants) interleaved order. In each of the four sessions of the pre-/post task, 366 images were acquired. Each of the four sessions of the learning task included 179 images.

The imaging data were preprocessed using SPM8 (Wellcome Department of Cognitive Neurology) for MATLAB (Mathworks, Natick, MA), FSL (http://www.fmrib.ox.ac.uk/fsl), and in-house scripts for the similarity analysis. Images were corrected for differences in slice acquisition timing and realigned to the mean image across all scans to correct for movement. Neither smoothing nor registration to standard space was performed, as all analyses were made in subject space. For group-level analyses of functional connectivity during the learning task, subject-level t-stats maps were smoothed and registered to MNI space (see below).

Regions of interest (ROIs). The hippocampus was defined anatomically for each participant using FSL's automatic subcortical segmentation protocol (FIRST). The hippocampus was segmented along its long axis by dividing the number of coronal slices in each hemisphere into three sections. The anterior third of the coronal slices was designated as anterior hippocampus, and the posterior third of the coronal slices was designated as posterior hippocampus[92]. We further divided the hippocampus a priori to the left and right hemisphere. We examined these four hippocampal ROIs (left/right by anterior/posterior), as previous findings on prior knowledge in the hippocampus do not coincide with respect to the specific locus of influence[30,33]. The left IFG and the AG were defined functionally, based on the group-level contrast of PK > n-PK in the PPI analysis detailed below. Then, for representational similarity analyses, we brought the peak voxel to each participant's native space and constructed a 12-mm sphere (10-mm sphere yielded similar results). Note that for the left AG, the peak voxel was located at the edge of the brain; thus, we chose the second peak (MNI coordinates: [−52, −48, 20]) for the representational similarity analysis.

Representational similarity analysis. For each subject, one GLM was constructed for the two pre-learning scans and one for the two post-learning scans. To model the response for each face in each session, the canonical hemodynamic response was convolved with the onset of the three presentations of the face in a session (time- derivative regressors were added, as well as a constant for each scan and a 128-s high-pass filter). This yielded a beta value for each stimulus in each of the four scans. We then converted these beta values into t statistics and averaged, for each stimulus, the two t-stats of the pre-learning scans, to obtain the multivoxel activity pattern before learning. The same was done for the two t-stats from the post-learning scans, to obtain the pattern of that face after learning. These multivoxel activity patterns were then correlated to obtain a similarity measure, and the Pearson's correlation coefficient was averaged (according to the specific analysis, as detailed below) and Fisher- transformed for statistical analysis. All the analysis steps after obtaining the t statistics were performed using a costume code in MATLAB R2018b (The MathWorks Inc), or in R (version 3.5.2[139]), where mentioned. More details can be found in the documentation at https://github.com/odedbein/SEL_public, where all the costume code is available.

We conducted three types of representational similarity analyses:

1. Memory-related pre-to-post similarity differences: We examined similarity differences that mediated explicit memory by computing for each participant the average similarity difference between pre- and post learning in each pair type (PK/n-PK), and within each memory outcome. That is, within each pair type, we averaged similarity differences for pairs that were remembered with high confidence in the subsequent memory test (high confidence included "certain" and "probably" responses; excluding low-confidence trials is a common practice used in fMRI studies to exclude guesses,[140] one participant with no high-confidence hits in the PK condition was removed from this analysis). We then compared this average to the average similarity for pairs that were forgotten (misses), within each pair type. To that end, differences in similarity from pre to post were entered into a repeated-measure ANOVA of pair type (PK/n-PK) by memory (remembered/forgotten). This ANOVA was followed up by two-tailed paired-sample *t* test, addressing simple effects.
2. Learning-related pre-to-post similarity differences: Prior to examining asymmetry in representational changes (see Introduction and below), it is important to know whether any change in similarity occurred from before to after learning. To that end, we computed for each participant, and for each pair type (PK and n-PK), the average similarity difference between pre- and post learning for pairs of items that appeared together during the learning task. As a control, we compared these similarity differences to the shuffled-pair baseline. The shuffled-pair baseline was obtained by pairing each A face with all other B faces of the same pair type, computing the similarity differences for each pair, and averaging across all pairs in each pair type. Note that this within-pair-type shuffling controls for any differences the B faces may have due to their appearance with famous versus novel pairs, since for each condition, all faces in the shuffled baseline appeared with A faces of the same pair type.
3. Asymmetry in representational changes: To assess whether cortical learning was asymmetric (see Introduction), for each pair that appeared together during learning, we subtracted the similarity of the A face post learning to that of the B face pre-learning, from the similarity of the B face post learning to the A face pre-learning[50]. This gave us a measure of how much more similar the B-face representation after learning became to that of the A face before learning, as compared to the extent to which the A face became similar to the B face. We then computed the average for each participant across all pairs per pair type (PK/n-PK). As control baseline, we further computed the same asymmetry index for shuffled pairs. Shuffled pairs were obtained by pairing each A face with all other B faces that did not appear with that A face during learning, but critically, appeared in the same pair type. We then averaged for each participant all shuffled pairs per pair type to obtain a baseline asymmetry index. Asymmetry was compared to 0 or to shuffled pairs using two-tailed paired-sample *t* tests.

Functional connectivity analysis. We conducted PPI analysis (SPM8 gPPI toolbox[61]) during the associative learning task, with the left anterior hippocampus ROI as a seed region (in pre- to post-learning similarity differences, this region showed separation for remembered PK pairs, but similarity for remembered n-PK pairs, see Results). Thus, for each participant, the times series of the left anterior hippocampus during the associative learning task was used as the physiological regressor. To reflect the nature of our task, the psychological regressors included a regressor for all pairs in each pair type, in each cycle of repetition (a total of 24 regressors, 12 repetitions by PK/n-PK pair type). The psychophysiological regressors were the interaction of each psychological regressor with the physiological regressor. As before, a high-pass filter of 128 s and constant scan regressors were added for each scan. We then computed, for each participant, the contrast of all repetitions of the PK pair type versus all repetitions of the n-PK pair type. Analyses were performed in each participant's native functional space. The resulting t maps were then smoothed (8-mm HFWM kernel) and registered to the MNI space for group-level analysis. At the group level, the PK versus n-PK contrast was compared to zero using a one-sample two-tailed *t* test. The resulting t map was thresholded at a voxel level of $P < 0.005$ due to low power in PPI designs[141], accounting for the reduced voxel-level threshold by maintaining a cluster-level threshold of $P < 0.05$[30] (cluster size >61 voxels, Monte Carlo simulations[142]).

Reproducibility. This is a single fMRI experiment. We did not repeat the experiment, and no replication attempts have been made to date.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Raw data and single-trial t-statistic maps that support the findings of this study are available online (https://osf.io/u2h3s/). A reporting summary for this article is available as a Supplementary Information file. Additional data are available from the corresponding author upon reasonable request. Source data are provided with this paper.

Code availability

The costume code that was used for data analysis in this study is available at https://github.com/odedbein/SEL_public.

Received: 21 January 2020; Accepted: 10 August 2020;
Published online: 14 September 2020

References

1. Alba, J. W. & Hasher, L. Is memory schematic? *Psychol. Bull.* **93**, 203–231 (1983).
2. Bein, O. et al. Delineating the effect of semantic congruency on episodic memory: the role of integration and relatedness. *PLoS ONE* **10**, e0115624 (2015).
3. Bransford, J. D. & Johnson, M. K. Contextual prerequisites for understanding: some investigations of comprehension and recall. *J. Verbal Learn. Verbal Behav.* **11**, 717–726 (1972).
4. Craik, F. I. M. & Tulving, E. Depth of processing and retention of words in episodic memory. *J. Exp. Psychol.* **104**, 268–294 (1975).
5. DeWitt, M. R., Knight, J. B., Hicks, J. L. & Ball, B. H. The effects of prior knowledge on the encoding of episodic contextual details. *Psychon. Bull. Rev.* **19**, 251–257 (2012).
6. Kole, J. A. & Healy, A. F. Using prior knowledge to minimize interference when learning large amounts of information. *Mem. Cogn.* **35**, 124–137 (2007).
7. Poppenk, J., Kohler, S. & Moscovitch, M. Revisiting the novelty effect: when familiarity, not novelty, enhances memory. *J. Exp. Psychol. Mem. Cogn.* **36**, 1321–1330 (2010).
8. Reder, L. M., Paynter, C., Diana, R. A., Ngiam, J. & Dickison, D. Experience is a double-edged sword: a computational model of the encoding/retrieval trade-off with familiarity. *Psychol. Learn. Motiv.* **48**, 271–312 (2007).
9. Reder, L. M. et al. Why it's easier to remember seeing a face we already know than one we don't: preexisting memory representations facilitate memory formation. *Psychol. Sci.* **24**, 363–372 (2013).
10. Fisher, R. P. & Craik, F. I. M. The effects of elaboration on recognition memory. *Mem. Cogn.* **8**, 400–404 (1980).
11. van Kesteren, M. T. R., Rijpkema, M., Ruiter, D. J. & Fernandez, G. Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. *J. Neurosci.* **30**, 15888–15894 (2010).
12. Bein, O., Trzewik, M. & Maril, A. The role of prior knowledge in incremental associative learning: an empirical and computational approach. *J. Mem. Lang.* **107**, 1–24 (2019).
13. Van Overschelde, J. P. & Healy, A. F. Learning of nondomain facts in high- and low-knowledge domains. *J. Exp. Psychol. Mem. Cogn.* **27**, 1160–1171 (2001).
14. Coutanche, M. N. & Thompson-Schill, S. L. Fast mapping rapidly integrates information into existing memory networks. *J. Exp. Psychol. Gen.* **143**, 2296–2303 (2014).
15. Ghosh, V. E. & Gilboa, A. What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia* **53**, 104–114 (2014).
16. Gilboa, A. & Marlatte, H. Neurobiology of schemas and schema-mediated memory. *Trends Cogn. Sci.* **21**, 618–631 (2017).
17. McClelland, J. L. Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *J. Exp. Psychol.* **142**, 1190–1210 (2013).
18. Tse, D. et al. Schemas and memory consolidation. *Science* **316**, 76–82 (2007).
19. Tse, D. et al. Schema-dependent gene activation and memory encoding in neocortex. *Science* **333**, 891–895 (2011).
20. Anderson, J. R. Retrieval of propositional information from long-term memory. *Cogn. Psychol.* **6**, 451–474 (1974).
21. Reder, L. M. & Anderson, J. R. The fan effect: new results and new theories. *J. Exp. Psychol. Gen.* **128**, 186–197 (1999).
22. Anderson, J. R. A spreading activation theory of memory. *J. Verbal Learn. Verbal Behav.* **22**, 261–295 (1983).
23. McClelland, J. L., McNaughton, B. L. & Oreilly, R. C. Why there are complementary learning-systems in the hippocampus and neocortex—insights from the success and failures of connectionist models of learning and memory. *Psychol. Rev.* **102**, 419–457 (1995).
24. Van Buuren, M. et al. Initial investigation of the effects of an experimentally learned schema on spatial associative memory in humans. *J. Neurosci.* **34**, 16662–16670 (2014).
25. van Kesteren, M. T. R., Fernandez, G., Norris, D. G. & Hermans, E. J. Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. *Proc. Natl Acad. Sci. USA* **107**, 7550–7555 (2010).
26. Wagner, I. C. et al. Schematic memory components converge within angular gyrus during retrieval. *eLife* **4**, 1–28 (2015).
27. Wang, S. H., Tse, D. & Morris, R. G. M. Anterior cingulate cortex in schema assimilation and expression. *Learn. Mem.* **19**, 315–318 (2012).
28. Amer, T., Giovanello, K. S., Nichol, D. R., Hasher, L. & Grady, C. L. Neural correlates of enhanced memory for meaningful associations with age. *Cereb. Cortex* **29**, 4568–4579 (2019).
29. Bein, O., Reggev, N. & Maril, A. Prior knowledge in fluences on hippocampus and medial prefrontal cortex interactions in subsequent memory. *Neuropsychologia* **64**, 320–330 (2014).
30. Brod, G., Lindenberger, U., Wagner, A. D. & Shing, Y. L. Knowledge acquisition during exam preparation improves memory and modulates memory formation. *J. Neurosci.* **36**, 8103–8111 (2016).
31. Kumaran, D., Summerfield, J. J., Hassabis, D. & Maguire, E. A. Tracking the emergence of conceptual knowledge during human decision making. *Neuron* **63**, 889–901 (2009).
32. Liu, Z. X., Grady, C. & Moscovitch, M. Effects of prior-knowledge on brain activation and connectivity during associative memory encoding. *Cereb. Cortex* **27**, 1991–2009 (2016).
33. Reggev, N., Bein, O. & Maril, A. Distinct neural suppression and encoding effects for conceptual novelty and familiarity. *J. Cogn. Neurosci.* **28**, 1455–1470 (2016).
34. Sommer, T. The emergence of knowledge and how it supports the memory for novel related information. *Cereb. Cortex* **27**, 1906–1921 (2017).
35. Preston, A. R. & Eichenbaum, H. Interplay of hippocampus and prefrontal cortex in memory. *Curr. Biol.* **23**, R764–R773 (2013).
36. van Kesteren, M. T. R., Ruiter, D. J., Fernández, G. & Henson, R. N. How schema and novelty augment memory formation. *Trends Neurosci.* **35**, 211–219 (2012).
37. Kuhl, B. A., Shah, A. T., DuBrow, S. & Wagner, A. D. Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. *Nat. Neurosci.* **13**, 501–506 (2010).
38. Baker, S. et al. The human dentate gyrus plays a necessary role in discriminating new memories. *Curr. Biol.* **26**, 2629–2634 (2016).
39. Bakker, A., Kirwan, C. B., Miller, M. & Stark, C. E. L. Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science* **319**, 1640–1642 (2008).
40. Chanales, A. J. H., Oza, A., Favila, S. E. & Kuhl, B. A. Overlap among spatial memories triggers repulsion of hippocampal representations article overlap among spatial memories triggers repulsion of hippocampal representations. *Curr. Biol.* **27**, 2307–2317 (2017).
41. Favila, S. E., Chanales, A. J. H. & Kuhl, B. A. Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nat. Commun.* **6**, 1–10 (2016).
42. Leutgeb, J. K., Leutgeb, S., Moser, M. & Moser, E. I. Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science* **315**, 961–966 (2007).
43. Maril, A. et al. Event congruency and episodic encoding: a developmental fMRI study. *Neuropsychologia* **49**, 3036–3045 (2011).
44. Staresina, B. P., Gray, J. C. & Davachi, L. Event congruency enhances episodic memory encoding through semantic elaboration and relational binding. *Cereb. Cortex* **19**, 1198–1207 (2009).
45. Badre, D., Poldrack, R. A., Paré-Blagoev, E. J., Insler, R. Z. & Wagner, A. D. Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron* **47**, 907–918 (2005).
46. Badre, D. & Wagner, A. D. Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia* **45**, 2883–2901 (2007).
47. Binder, J. R. & Desai, R. H. The neurobiology of semantic memory. *Trends Cogn. Sci.* **15**, 527–536 (2011).
48. Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K. & Farah, M. J. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proc. Natl Acad. Sci. USA* **94**, 14792–14797 (1997).
49. Kim, G., Norman, K. A. & Turk-Browne, N. B. Neural differentiation of incorrectly predicted memories. *J. Neurosci.* **37**, 2022–2031 (2017).
50. Schapiro, A. C., Kustner, L. V. & Turk-Browne, N. B. Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr. Biol.* **22**, 1622–1627 (2012).
51. Schlichting, M. L., Mumford, J. A. & Preston, A. R. Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nat. Commun.* **6**, 1–10 (2015).
52. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis—connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 1–28 (2008).
53. Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional brain mapping. *Proc. Natl Acad. Sci. USA* **103**, 3863–3868 (2006).

54. Davachi, L. Item, context and relational episodic encoding in humans. *Curr. Opin. Neurobiol.* **16**, 693–700 (2006).
55. Diana, R. A., Yonelinas, A. P. & Ranganath, C. Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends Cogn. Sci.* **11**, 379–386 (2007).
56. Eichenbaum, H., Yonelinas, A. P. & Ranganath, C. The medial temporal lobe and recognition memory. *Annu. Rev. Neurosci.* **30**, 123–152 (2007).
57. Squire, L. R. Memory systems of the brain: a brief history and current perspective. *Neurobiol. Learn. Mem.* **82**, 171–177 (2004).
58. 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).
59. Mumford, J. A., Davis, T. & Poldrack, R. A. The impact of study design on pattern estimation for single-trial multivariate pattern analysis. *Neuroimage* **103**, 130–138 (2014).
60. Gobbini, M. I. & Haxby, J. V. Neural systems for recognition of familiar faces. *Neuropsychologia* **45**, 32–41 (2007).
61. McLaren, D. G., Ries, M. L., Xu, G. F. & Johnson, S. C. A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. *Neuroimage* **61**, 1277–1286 (2012).
62. Binder, J. R., Desai, R. H., Graves, W. W. & Conant, L. L. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* **19**, 2767–2796 (2009).
63. Natu, V. & O'Toole, A. J. The neural processing of familiar and unfamiliar faces: a review and synopsis. *Br. J. Psychol.* **102**, 726–747 (2011).
64. Eichenbaum, H. Prefrontal–hippocampal interactions in episodic memory. *Nat. Rev. Neurosci.* **18**, 547–558 (2017).
65. Sharon, T., Moscovitch, M. & Gilboa, A. Rapid neocortical acquisition of long-term arbitrary associations independent of the hippocampus. *Proc. Natl Acad. Sci. USA* **108**, 1146–1151 (2011).
66. Bein, O., Reggev, N. & Tompary, A. Working with schemas, predicting with schemas. *J. Neurosci.* **38**, 1608–1610 (2018).
67. van der Linden, M., Berkers, R. M. W. J., Morris, R. G. M. & Fernández, G. Angular gyrus involvement at encoding and retrieval is associated with durable but less specific memories. *J. Neurosci.* **37**, 9474–9485 (2017).
68. Coutanche, M. N. Distinguishing multi-voxel patterns and mean activation: why, how, and what does it tell us? *Cogn. Affect. Behav. Neurosci.* **13**, 667–673 (2013).
69. Davis, T. & Poldrack, R. A. Measuring neural representations with fMRI: practices and pitfalls. *Ann. N. Y. Acad. Sci.* **1296**, 108–134 (2013).
70. LaRocque, K. F. et al. Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *J. Neurosci.* **33**, 5466–5474 (2013).
71. Long, D. L. & Prat, C. S. Memory for star trek: the role of prior knowledge in recognition revisited. *J. Exp. Psychol. Mem. Cogn.* **28**, 1073–1082 (2002).
72. Bakker, A., Kirwan, C. B., Miller, M. & Stark, C. E. L. Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science* **319**, 1640–1642 (2009).
73. Simons, J. S. & Spiers, H. J. Prefrontal and medial temporal lobe interactions in long-term memory. *Nat. Rev. Neurosci.* **4**, 637–648 (2003).
74. Eichenbaum, H. Memory: organization and control. *Annu. Rev. Psychol.* **68**, 19–45 (2017).
75. 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. *J. Cogn. Neurosci.* **27**, 96–111 (2015).
76. Guise, K. G. & Shapiro, M. L. Medial prefrontal cortex reduces memory interference by modifying hippocampal encoding. *Neuron* **94**, 183–192 (2017).
77. Rajasethupathy, P. et al. Projections from neocortex mediate top-down control of memory retrieval. *Nature* **526**, 653–659 (2015).
78. Shimamura, A. P., Jurica, P. J., Mangels, J. A., Gershberg, F. B. & Knight, R. T. Susceptibility to memory interference effects following frontal lobe damage: findings from tests of paired-associate learning. *J. Cogn. Neurosci.* **7**, 144–152 (1995).
79. Kafkas, A. & Montaldi, D. How do memory systems detect and respond to novelty? *Neurosci. Lett.* **680**, 60–68 (2018).
80. Rangel-Gomez, M. & Meeter, M. Neurotransmitters and novelty: a systematic review. *J. Psychopharmacol.* **30**, 3–12 (2016).
81. Duncan, K. D. & Schlichting, M. L. Hippocampal representations as a function of time, subregion, and brain state. *Neurobiol. Learn. Mem.* **153**, 40–56 (2018).
82. Giocomo, L. M. & Hasselmo, M. E. Neuromodulation by glutamate and acetylcholine can change circuit dynamics by regulating the relative influence of afferent input and excitatory feedback. *Mol. Neurobiol.* **36**, 184–200 (2007).
83. Hasselmo, M. E., Bodelón, C. & Wyble, B. P. A proposed function for hippocampal theta rhythm: separate phases of encoding and retrieval enhance reversal of prior learning. *Neural Comput.* **14**, 793–817 (2002).
84. Hasselmo, M. E., Wyble, B. P. & Wallenstein, G. V. Encoding and retrieval of episodic memories: role of cholinergic and GABAergic modulation in the hippocampus. *Hippocampus* **6**, 693–708 (1996).
85. Lisman, J. E. & Grace, A. A. The hippocampal-VTA loop: controlling the entry of information into long-term memory. *Neuron* **46**, 703–713 (2005).
86. Meeter, M., Murre, J. M. J. & Talamini, L. M. Mode shifting between storage and recall based on novelty detection in oscillating hippocampal circuits. *Hippocampus* **14**, 722–741 (2004).
87. Giocomo, L. M. & Hasselmo, M. E. Computation by oscillations: Implications of experimental data for theoretical models of grid cells. *Hippocampus* **18**, 1186–1199 (2008).
88. Hansen, N. & Manahan-Vaughan, D. Locus coeruleus stimulation facilitates long-term depression in the dentate gyrus that requires activation of β-adrenergic receptors. *Cereb. Cortex* **25**, 1889–1896 (2015).
89. Harley, C. W. Norepinephrine and the dentate gyrus. *Prog. Brain Res.* **163**, 299–318 (2007).
90. Brunec, I. K. et al. Multiple scales of representation along the hippocampal anteroposterior axis in humans. *Curr. Biol.* **28**, 2129–2135 (2018).
91. Poppenk, J., Evensmoen, H. R., Moscovitch, M. & Nadel, L. Long-axis specialization of the human hippocampus. *Trends Cogn. Sci.* **17**, 230–240 (2013).
92. Tompary, A. & Davachi, L. Consolidation promotes the emergence of representational overlap in the hippocampus and medial prefrontal cortex. *Neuron* **96**, 228–241 (2017).
93. Burwell, D. The parahippocampal region: corticocortical connectivity. *Ann. N. Y. Acad. Sci.* **911**, 25–42 (2000).
94. Suzuki, W. A. & Amaral, D. G. Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents 4025. *J. Comp. Neurol.* **350**, 497–533 (1994).
95. Barense, M. D. et al. Intact memory for irrelevant information impairs perception in amnesia. *Neuron* **75**, 157–167 (2012).
96. Clarke, A. & Tyler, L. K. Object-specific semantic coding in human perirhinal cortex. *J. Neurosci.* **34**, 4766–4775 (2014).
97. Davachi, L., Mitchell, J. P. & Wagner, A. D. Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. *Proc. Natl Acad. Sci. USA* **100**, 2157–2162 (2003).
98. Staresina, B. P. & Davachi, L. Selective and shared contributions of the hippocampus and perirhinal cortex to episodic item and associative encoding. *J. Cogn. Neurosci.* **20**, 1478–1489 (2008).
99. Staresina, B. P., Duncan, K. & Davachi, L. Perirhinal and parahippocampal cortices differentially contribute to later recollection of object- and scene-related event details. *J. Neurosci.* **31**, 8739–8747 (2011).
100. Brown, M. W. & Aggleton, J. P. Recognition memory: what are the roles of the perirhinal cortex and hippocampus? *Nat. Rev. Neurosci.* **2**, 51–61 (2001).
101. Ranganath, C. & Ritchey, M. Two cortical systems for memory-guided behaviour. *Nat. Rev. Neurosci.* **13**, 713–726 (2012).
102. Zeithamova, D. & Preston, A. R. Temporal proximity promotes integration of overlapping events. *J. Cogn. Neurosci.* **29**, 1311–1323 (2017).
103. Poppenk, J. & Norman, K. A. Mechanisms supporting superior source memory for familiar items: a multi-voxel pattern analysis study. *Neuropsychologia* **50**, 3015–3026 (2012).
104. Duncan, K., Sadanand, A. & Davachi, L. Memory's penumbra: episodic memory decisions induce lingering mnemonic biases. *Science* **337**, 485–487 (2012).
105. Hasselmo, M. E. & Stern, C. E. Theta rhythm and the encoding and retrieval of space and time. *Neuroimage* **85**, 656–666 (2014).
106. Knierim, J. J. & Neunuebel, J. P. Tracking the flow of hippocampal computation: pattern separation, pattern completion, and attractor dynamics. *Neurobiol. Learn. Mem.* **129**, 38–49 (2016).
107. Marr, D. Simple memory: a theory for archicortex. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **262**, 23–81 (1971).
108. O'Reilly, R. C. & McClelland, J. L. Hippocampal conjunctive encoding, storage, and recall: avoiding a trade-off. *Hippocampus* **4**, 661–682 (1994).
109. Treves, A. & Rolls, E. T. Computational analysis of the role of hippocampus in memory. *Hippocampus* **4**, 374–391 (1994).
110. Bosch, S. E., Jehee, J. F. M., Fernandez, G. & Doeller, C. F. Reinstatement of associative memories in early visual cortex is signaled by the hippocampus. *J. Neurosci.* **34**, 7493–7500 (2014).
111. Danker, J. F., Tompary, A. & Davachi, L. Trial-by-trial hippocampal encoding activation predicts the fidelity of cortical reinstatement during subsequent retrieval. *Cereb. Cortex* **27**, 3515–3524 (2017).
112. Ritchey, M., Wing, E. A., Labar, K. S. & Cabeza, R. Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. *Cereb. Cortex* **23**, 2818–2828 (2013).
113. Staresina, B. P., Henson, R. N. A., Kriegeskorte, N. & Alink, A. Episodic reinstatement in the medial temporal lobe. *J. Neurosci.* **32**, 18150–18156 (2012).
114. Bein, O., Duncan, K. & Davachi, L. Mnemonic prediction errors bias hippocampal states Oded. *Nat. Commun.* **11**, 1–11 (2020).
115. Colgin, L. L. Rhythms of the hippocampal network. *Nat. Rev. Neurosci.* **17**, 239–249 (2016).

116. Hasselmo, M. E., Schnell, E. & Barkai, E. Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *J. Neurosci.* **15**, 5249–5262 (1995).
117. Ahissar, M. & Hochstein, S. The reverse hierarchy theory of visual perceptual learning. *Trends Cogn. Sci.* **8**, 457–464 (2004).
118. Brodt, S. et al. Fast track to the neocortex: a memory engram in the posterior parietal cortex. *Science* **362**, 1045–1048 (2018).
119. Brodt, S. et al. Rapid and independent memory formation in the parietal cortex. *Proc. Natl Acad. Sci. USA* **113**, 13251–13256 (2016).
120. Hebscher, M., Wing, E., Ryan, J. & Gilboa, A. Rapid cortical plasticity supports long-term memory formation. *Trends Cogn. Sci.* **23**, 989–1002 (2019).
121. Heeger, D. J. Theory of cortical function. *Proc. Natl Acad. Sci. USA* **114**, 1773–1782 (2017).
122. Moscovitch, M., Cabeza, R., Winocur, G. & Nadel, L. Episodic memory and beyond: the hippocampus and neocortex in transformation. *Annu. Rev. Psychol.* **67**, 105–134 (2016).
123. Popov, V., Zhang, Q., Koch, G. E., Calloway, R. C. & Coutanche, M. N. Semantic knowledge influences whether novel episodic associations are represented symmetrically or asymmetrically. *Mem. Cogn.* **47**, 1567–1581 (2019).
124. Bellezza, F. S. & Buck, D. K. Expert Knowledge as mnemonic cues. *Appl. Cogn. Psychol.* **2**, 147–162 (1988).
125. Schacter, D. L., Dobbins, I. G. & Schnyer, D. M. Specificity of priming: a cognitive neuroscience perspective. *Nat. Rev. Neurosci.* **5**, 853–862 (2004).
126. Tulving, E. & Schacter, D. L. Priming and human memory systems. *Science* **247**, 302–306 (1990).
127. Reber, P. J. The neural basis of implicit learning and memory: a review of neuropsychological and neuroimaging research. *Neuropsychologia* **51**, 2026–2042 (2013).
128. Hayama, H. R., Vilberg, K. L. & Rugg, M. D. Overlap between the neural correlates of cued recall and source memory: evidence for a generic recollection network? *J. Cogn. Neurosci.* **24**, 1127–1137 (2012).
129. Rugg, M. D. & Vilberg, K. L. Brain networks underlying episodic memory retrieval. *Curr. Opin. Neurobiol.* **23**, 255–260 (2013).
130. Thakral, P. P., Wang, T. H. & Rugg, M. D. Decoding the content of recollection within the core recollection network and beyond. *Cortex* **91**, 101–113 (2017).
131. Atienza, M., Crespo-Garcia, M. & Cantero, J. L. Semantic congruence enhances memory of episodic associations: role of theta oscillations. *J. Cogn. Neurosci.* **23**, 75–90 (2011).
132. Verosky, S. C., Todorov, A. & Turk-Browne, N. B. Representations of individuals in ventral temporal cortex defined by faces and biographies. *Neuropsychologia* **51**, 2100–2108 (2013).
133. Peelen, M. V. & Caramazza, A. Conceptual object representations in human anterior temporal cortex. *J. Neurosci.* **32**, 15728–15736 (2012).
134. Thierry, G., Martin, C. D., Downing, P. & Pegna, A. J. Controlling for interstimulus perceptual variance abolishes N170 face selectivity. *Nat. Neurosci.* **10**, 505–511 (2007).
135. Dale, A. M. Optimal experimental design for event-related fMRI. *Hum. Brain Mapp.* **8**, 109–114 (1999).
136. Bentin, S. & Deouell, L. Y. Structural encoding and identification in face processing: ERP evidence for separate mechanisms. *Cogn. Neuropsychol.* **17**, 35–54 (2000).
137. Neumann, M. F. & Schweinberger, S. R. N250r and N400 ERP correlates of immediate famous face repetition are independent of perceptual load. *Brain Res.* **1239**, 181–190 (2008).
138. Tacikowski, P., Jednoróg, K., Marchewka, A. & Nowicka, A. How multiple repetitions influence the processing of self-, famous and unknown names and faces: an ERP study. *Int. J. Psychophysiol.* **79**, 219–230 (2011).
139. R core team. R: a language and environment for statistical computing. https://www.r-project.org/ (R Foundation for Statistical Computing, Vienna, Austria, 2018).
140. Wagner, A. D. et al. Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science* **281**, 1188–1191 (1998).
141. O'Reilly, J. X., Woolrich, M. W., Behrens, T. E. J., Smith, S. M. & Johansen-Berg, H. Tools of the trade: psychophysiological interactions and functional connectivity. *Soc. Cogn. Affect. Neurosci.* **7**, 604–609 (2012).
142. Slotnick, S. D., Moo, L. R., Segal, J. B. & Hart, J. Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Cogn. Brain Res.* **17**, 75–82 (2003).

Acknowledgements

This work was supported by Israel Science Foundation (150/16, to A.M.). We thank Haim Cohen and Alexa Tompary for facilitating the fMRI analyses and Maayan Trzewik for her help in data collection. We further thank Ram Frost, Alexa Tompary, David Clewett, Tarek Amer, Catherine Hartley, and Katherine Nussbaum for their insightful comments and suggestions in preparing the paper. We thank Danielle Babitz for her help in preparing the photos used in the paper.

Author contributions

O.B., N.R., and A.M. designed the experiment; O.B. collected and analyzed the data; A.M. provided guidance; O.B. wrote the initial draft of the paper; O.B., N.R., and A.M. revised and edited the paper; A.M. acquired funding.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-18364-1.

Correspondence and requests for materials should be addressed to A.M.

Peer review information *Nature Communication* thanks the anonymous reviewers for their contributions to the peer review of this work. Peer review reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.