Research article

Method of loci training yields unique neural representations that support effective memory encoding

Jingyuan Ren[1], Boris N. Konrad[1], Yannan Zhu[1], Fan Li[1,2], Michael Czisch[3], Martin Dresler[1,#] & Isabella C. Wagner[4,5, #, *]

[1] Donders Institute for Brain, Cognition, and Behaviour, Radboud University Medical Center, 6525 EN Nijmegen, Netherlands; [2] School of Biomedical Engineering, Faculty of Medicine, Dalian University of Technology, Dalian 116024, China; [3] Max Planck Institute of Psychiatry, 80804 Munich, Germany; [4] Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, 1010 Vienna, Austria; [5] Vienna Cognitive Science Hub, University of Vienna, 1090 Vienna, Austria.

Shared last authors

* **Lead contacts**: Isabella C. Wagner, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria, Phone: +43 (0)1 4277 47151, Email: isabella.wagner@univie.ac.at

Abstract

The method of loci is a technique to effectively boost memory, but its impact on the underlying neural representations is poorly understood. Here, we used functional magnetic resonance imaging and representational similarity analysis to compare the neural representations of memory athletes ranked among the world's top 50 in memory sports to those of mnemonics-naïve controls. In a second study, mnemonics-naïve individuals underwent a 6-week-long memory training, working memory training, or no intervention. Results showed distinct neural representations in the prefrontal cortex, inferior temporal, and posterior parietal regions as memory athletes and the memory training group studied novel content. Neural representations were also distinct between these experienced individuals, which was related to better memory performance after 4 months. Our findings highlight how extensive memory training affects neocortical memory engrams. We suggest that the method of loci may bolster memory uniqueness within one's "memory palace", setting the stage for exceptional memory performance.

Keywords

Memory training, method of loci, functional magnetic resonance imaging (fMRI), representational similarity analysis (RSA), lateral prefrontal cortex

Introduction

Mnemonic techniques can substantially boost memory performance. One of the most commonly used techniques is the so-called method of loci relying on mental navigation along well-known spatial routes (Yates, 1966). To-be-studied material is mentally placed at salient locations along the imagined route and can then be recalled by "retracing" the route, "picking up" the material where it was previously "dropped off". Successful application of the method of loci requires extensive training and is often used by experts, or "memory athletes", who demonstrate their superior abilities in memorizing and accurately reproducing tremendous amounts of arbitrary information in competitions such as the World Memory Championships. Previously, our group showed that also mnemonic-naïve individuals can enhance their memory performance through multiple-week training (Dresler et al., 2017). The training effects were accompanied by changes in neural activation and connectivity of brain regions typically engaged in memory and visuospatial processing, including the hippocampus, posterior medial, and prefrontal regions (Wagner et al., 2021). Parallel to levels of activation and connectivity, superior memory might rely on distributed activation patterns that can be assessed with representational similarity analysis (RSA, Kriegeskorte et al., 2008). RSA quantifies the spatial alignment of activation patterns and dissimilar patterns are thought to code for distinct informational content. For instance, Liu and colleagues (2022) demonstrated that method of loci training over the course of five days was associated with lower neural pattern similarity in hippocampal subfields during memory encoding (hence, distinct hippocampal patterns). This effect was specific to the studied material that was mentally placed at close-by locations along the imagined route and was coupled to accurate temporal order memory. Together, this highlights that short-term training with the method of loci disambiguates hippocampal representations, but the question remains of how extended, multiple-week training transforms the neural representations that underlie superior memory abilities as seen in memory athletes.

The processes that enable superior memory through method of loci training may also impact the similarity of neural patterns between individuals. This is grounded upon recent work demonstrating that the degree of neural pattern similarity to an average, group-based neural pattern during associative encoding predicted subsequent memory ability across individuals (Sheng et al., 2023). Put differently, the more the individuals resembled the group in their neural representations, the better their memory was. Increased neural pattern similarity between individuals was also reported during successful scene encoding (Koch et al., 2020) and associative retrieval (Xiao et al., 2020). Such "neural alignment" between individuals might stem from the processing of similar content. For instance, individuals display neural alignment when watching the same video (Chen et al., 2017a), or when sharing beliefs about the same story (Yeshurun et al., 2017). In educational settings, neural alignment between students is tied to knowledge acquisition of abstract physics concepts (Mason & Just, 2016) and indicates effective information transmission in the classroom (Nguyen et al., 2022). Taking this a step further, Meshulam and colleagues (2021) not only showed neural alignment between learners after acquiring new conceptual knowledge but also reported neural alignment between learners and experts that was predictive of learning outcomes. Whether superior memory after method of loci training is coupled to neural alignment between individuals (and, specifically, between trainees and memory athletes, akin to learners and experts) is currently unclear. Furthermore, the notion of neural alignment appears difficult to reconcile with the finding of distinct neural representations after method of loci training (Liu et al., 2022), raising the possibility that neural alignment might be less beneficial when memory specificity and flexible order recognition are at stake.

We addressed these knowledge gaps in two separate studies (for an overview, see Fig. 1). In the first study (the "athlete study"), we included individuals who were experts in using the method of loci (hereafter referred to as "memory athletes") and who were ranked among the world's top 50 in memory sports (n = 17). These were compared to a control group that was closely matched for age, sex, handedness, and intelligence (n = 16; see Table S1 for demographic details). To assess neural representations during memory encoding and retrieval, participants underwent functional magnetic resonance imaging (fMRI) while studying novel words (word list encoding task) and were subsequently tested for their ability to recognize

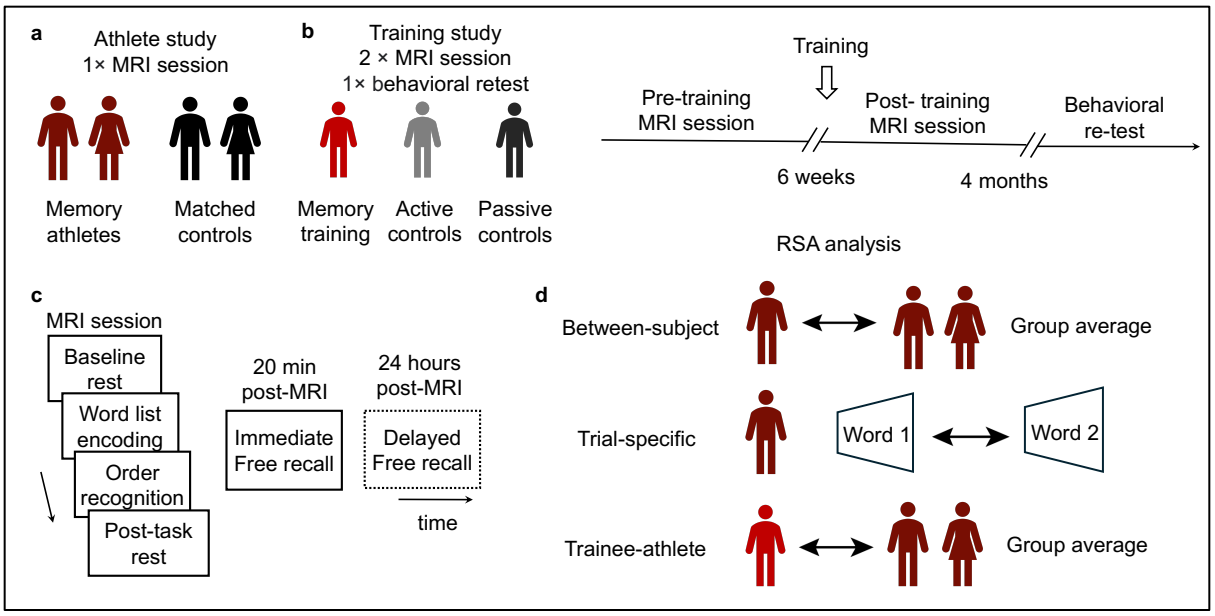

Fig. 1. Overview of athlete and training studies, and representational similarity analysis. (**a**) We examined memory athletes (*n* = 17) and compared them to matched controls (*n* = 16) in a single MRI session (athlete study). (**b**) Individuals in the training study were assigned to one of three groups (memory training group, *n* = 17, active controls, *n* = 16, passive controls, *n* = 17). They completed two MRI sessions prior and after the 6-week-long interval (pre- and post-training sessions), and a behavioural re-test after 4 months. (**c**) General structure of all MRI sessions: baseline and post-task rest periods (not discussed further), word list encoding and temporal order recognition tasks (10 min each), immediate (20 min post-MRI) and delayed (24 hours post-MRI) free recall (dashed frame indicates that the delayed test was only included in the training study). (**d**) We conducted three types of representational similarity analysis (RSA). Between-subject RSA: this assessed the neural pattern similarity between individuals within the same experimental group. Trial-specific RSA: this examined neural pattern similarity across the studied words but within each individual. Trainee-athlete RSA: this evaluated the degree of neural pattern similarity between trainees and memory athletes. Note that the schematic shows memory athletes and individuals of the memory training group (the "trainees") only, but the same analysis was performed for all remaining groups as well.

their temporal order (temporal order recognition task). We specifically opted for these tasks as they are commonly used during memory championships, and because we reasoned that the particular strength of the method of loci lies in the learning and recalling of ordered sequences (due to mental navigation through the "memory palace"). Memory performance was assessed through a free recall test after fMRI. In the second study (the "memory training study"), we enrolled initially mnemonics-naïve individuals who completed a 6-week method of loci training (memory training group, *n* = 17). These individuals were compared to two control groups that completed a 6-week working memory training (active controls, *n* = 16) or no intervention (passive controls, *n* = 17). All participants completed the abovementioned memory tasks during fMRI before and after the 6-week period to determine any (training-related) changes in neural representations. Memory performance was assessed in a free recall test post-fMRI and after 24 hours, as well as in a re-test after four months.

Since the method of loci involves the active formation of unique, salient associations (Liu et al., 2022), we hypothesized distinct neural representations for the verbal material that was studied during the word list encoding task. This effect should be specific to individuals who completed the method of loci training (thus, memory athletes and the memory training group, or the so-called "trainees"). Our main analysis was focused on the training-related changes in neural representations between individuals. We reasoned that the distinct neural representations on the level of the studied material should impact the degree of neural alignment between those who completed the method of loci training. This allowed us to directly test whether trainee-athlete neural alignment during encoding is beneficial when memory specificity and temporal order recognition are at stake. Lastly, we leveraged the data from the recognition task. This allowed us to explore

whether method of loci training has comparable effects on neural representations when individuals retrieve the previously learned word-location associations. Overall, our approach provides first insights into how extended memory training transforms the neural representations that give rise to superior memory.

Results

Memory training is associated with distinct neural representations between experienced individuals during word list encoding

Building on our previous work that showed increased free recall performance after method of loci training (Dresler et al., 2017) as well as enhanced durable memory formation (Wagner et al., 2021), we here revisited the fMRI data to investigate potential changes in the underlying neural representations that might arise after memory training. Because successful memory formation after shorter-term method of loci training was shown to be associated with distinct neural representations of the specific study material (Liu et al., 2022), we hypothesized that this would impact the degree of neural alignment between experienced individuals. In other words, method of loci training should lead to distinct neural representations (thus, lower neural alignment) between memory athletes and between individuals of the memory training group after training while they studied novel words (word list encoding task; Fig. 2b). During the task, memory athletes and individuals of the memory training group (post-training) were asked to use the method of loci (they were asked to mentally navigate through their memory palace and to "place" the studied words at the specific location). We quantified the degree of neural alignment by means of representational similarity analysis (RSA) and focused on brain regions that we previously found to be associated with word list encoding following method of loci training, such as the lateral prefrontal cortex and the superior frontal gyrus (Wagner et al., 2021). For each individual and region-of-interest (ROIs), we extracted the average neural pattern during word list encoding and correlated it with the neural pattern of all other individuals of the same experimental group (e.g., memory athletes × memory athletes, and memory training group × memory training group, separately for pre- and post-training). This yielded the magnitude of neural pattern similarity per group and time point (Fig. 2a, see section "Between-subject RSA" in the Methods).

First, we investigated data from the athlete study. Results revealed significantly lower neural pattern similarity between memory athletes compared to matched controls in both the left lateral prefrontal cortex (separate independent-sample *t*-tests, memory athletes, $n = 17$, matched controls, $n = 16$, $t(31) = 2.690$, $p = 0.011$, Cohen's $d = 0.937$; Fig. 2d) and the left superior frontal gyrus during word list encoding ($t(31) = 4.350$, $p < 0.001$, Cohen's $d = 1.517$; Fig. 2d). To test for effects beyond these pre-defined ROIs, we repeated our analysis but instead moved a spherical searchlight throughout the entire brain volume (see section "Searchlight-based RSA" in the Methods). This confirmed the abovementioned results and further showed significantly lower neural pattern similarity in the left inferior temporal cortex, the superior parietal and precentral gyrus in memory athletes compared to matched controls (independent-sample *t*-test, same samples as above, contrasts word list encoding > baseline in memory athletes < matched controls, $p < 0.05$, family-wise error (FWE) corrected at cluster-level using a cluster-defining threshold of $p < 0.001$, critical cluster size = 36 voxels, Fig. 2e, Table S2). Thus, in line with our expectations, we found distinct neural representations (lower neural alignment) between memory athletes compared to matched controls in brain regions related to successful memory encoding and visuospatial processing when studying novel words.

Next, we turned to the data of the training study. Similar to the findings above, we observed significantly lower neural pattern similarity between the individuals of the memory training group after training (compared to the pre-training baseline) in the left lateral prefrontal cortex (separate paired-sample *t*-tests, $n = 17$, $t(16) = 2.200$, $p = 0.043$, Cohen's $d = 0.533$, Fig. 2f) and the left superior frontal gyrus ($t(16) = 7.240$, $p < 0.001$, Cohen's $d = 1.756$, Fig. 2f). This effect was specific to the memory training group after training but was not present in the active or passive control groups (Fig. S1). Whole-brain searchlight analysis confirmed these findings and further revealed lower neural pattern similarity in individuals of the memory

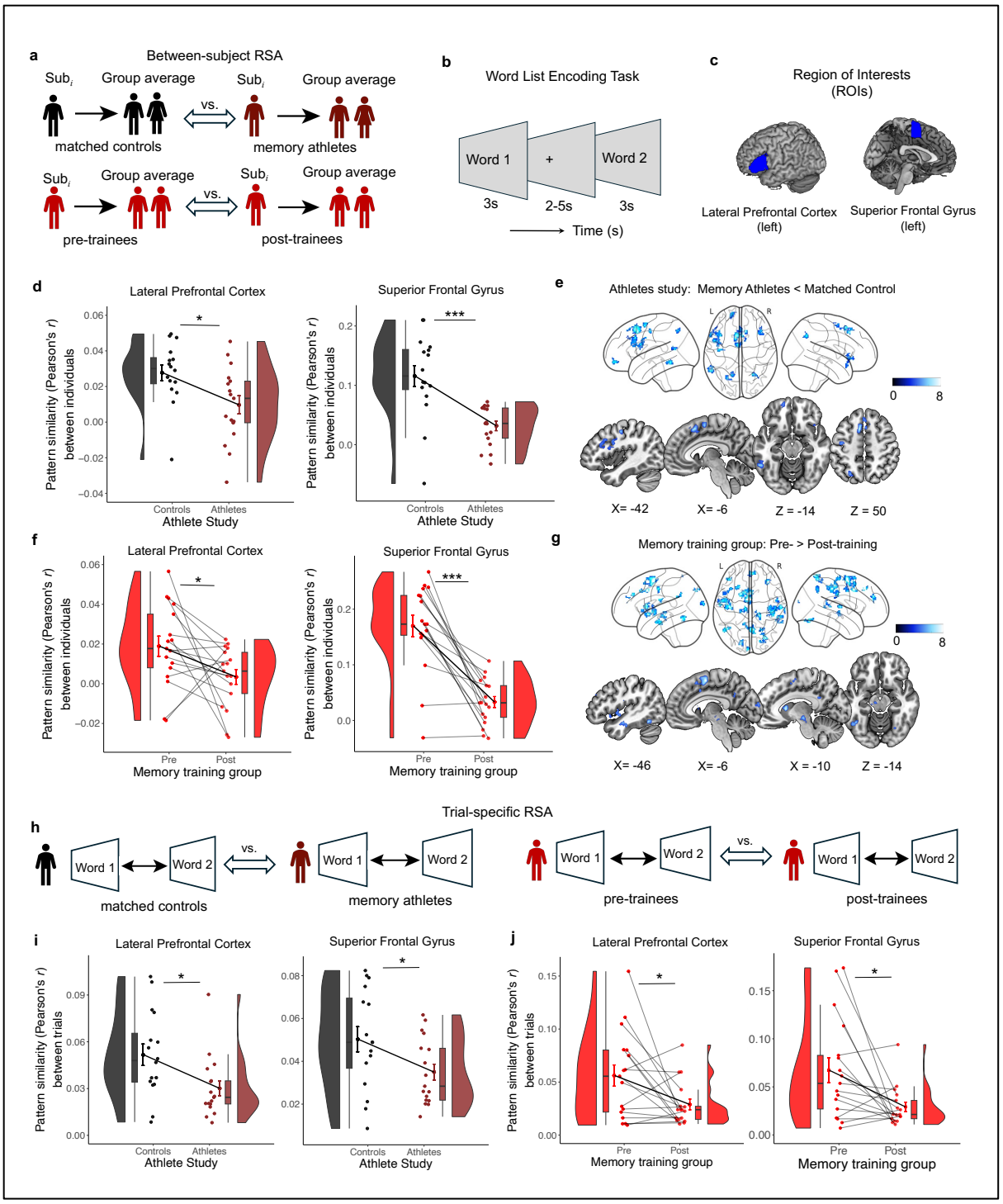

Fig. 2. Method of loci training yields distinct neural representations between experienced individuals during word list encoding. (**a**) Detailed schematic of the between-subject RSA: We compared the individual-to-group neural pattern similarity between memory athletes (n = 17) and matched controls (n = 16) during word list encoding, and between pre- and post-training within the memory training group (n = 17, see Methods, and see supplementary materials for control analyses). (**b**) Schematic of the word list encoding task. (**c**) Regions of interest (ROIs) that were used for all analyses pertaining to word list encoding. (**d, f**) ROI results of the between-subject RSA (control analyses for the active and passive control groups are shown in Fig. S1). (**e, g**) Searchlight results of the between-subject RSA. Results were thresholded at $p < 0.05$, family-wise error (FWE)-corrected at cluster-level, using a cluster-defining threshold of $p < 0.001$ (Tables S2-3). (**h**) Detailed schematic of the trial-specific RSA: We calculated neural pattern similarity on a trial-by-trial basis for each participant and compared memory athletes to matched controls, and pre- to

post-training time points in the memory training group (same *n* as above). (**i, j**) ROI results of the trial-specific RSA (control analyses for the active and passive control groups are shown in Fig. S3). (**d, f, i, j**) Data points represent individual participants (athlete study, $n = 33$, memory training group, $n = 17$). Boxplots display the median and upper/lower quartile, with whiskers extending to the most extreme data points within 1.5 interquartile ranges above/below the quartiles. The data points with error bars show the mean ± standard error of the mean (S.E.M.). Data distributions are based on the respective probability density function. All ROI-based results survived multiple comparison corrections using the false discovery rate (FDR) with the Benjamini–Hochberg procedure set to $q < 0.05$. ***$p < 0.001$; *$p < 0.05$; ns, not significant.

training group after training (compared to the pre-training baseline) in the left inferior temporal cortex, caudate nucleus, pre- and postcentral gyri, the superior temporal and parietal gyrus, among others (paired-sample *t*-test, $n = 17$, contrast word list encoding > baseline during pre-training > post-training, $p < 0.05$, FWE-corrected, cluster-defining threshold $p < 0.001$, critical cluster size = 30 voxels, Fig 2g, Table S3). Neural pattern similarity in the memory training group was also lower as compared to the active and passive controls (Fig. S2). Overall, studying word lists while utilizing the method of loci was coupled to distinct neural representations between memory athletes, and between individuals of the memory training group after training.

Experienced individuals hold unique neural representations associated with the studied words

To verify that the finding of distinct neural representations between experienced individuals was related to distinct neural representations associated with each studied word (Liu et al., 2022), we focused on the same *a priori* ROIs as above and quantified the average neural pattern similarity across all studied words per participant (see section "Trial-specific RSA" in the Methods and Fig. 2h).

Results indeed revealed significantly lower neural pattern similarity associated with the studied words in memory athletes compared to matched controls in both the left lateral prefrontal cortex (separate independent-sample *t*-tests, memory athletes, $n = 17$, matched controls, $n = 16$, $t(31) = 2.590$, $p = 0.014$, Cohen's $d = 0.904$, Fig. 2i) and the left superior frontal gyrus ($t(31) = 2.230$, $p = 0.033$, Cohen's $d = 0.778$; Fig. 2i). Returning to the data from the training study, we observed significantly lower neural pattern similarity associated with the studied words in the memory training group after training (compared to the pre-training baseline) in both the left lateral prefrontal cortex (separate paired-sample *t*-tests, $n = 17$, $t(16) = 2.530$, $p = 0.022$, Cohen's $d = 0.613$; Fig. 2j) and the left superior frontal gyrus ($t(16) = 2.800$, $p = 0.013$, Cohen's $d = 0.679$; Fig. 2j). This effect was only present in the memory training group after training, but not in the active or passive controls (Fig. S3). Thus, the distinct neural representations between individuals who were proficient in the method of loci appeared driven by distinct neural representations associated with each studied word.

Distinct neural representations between trainees and memory athletes during word list encoding

Previous work showed that the degree of neural alignment between learners and experts was predictive of learning outcomes (Meshulam et al., 2022), and we were eager to test whether similar mechanisms are at play when it comes to expertise with the method of loci. So far, we have shown unique neural representations for the studied material, distinct neural representations between individuals of the memory training group after training (thus, in the so-called "trainees"), and between memory athletes. Based on these results, we expected distinct neural representations between trainees and memory athletes, which would indicate that superior memory relies on neural processes that are specific to those who are proficient in utilizing the method of loci. To quantify the degree of trainee-athlete similarity in neural representations, we extracted the average neural patterns during word list encoding for each participant of the memory training group. Analyses were performed using the abovementioned *a priori*-defined ROIs. We then correlated the neural patterns with those of each memory athlete, and, separately, with those of each individual of the matched control group. This allowed us to compare the average neural pattern similarity

between *time points* (pre- vs. post-training) and *groups* (neural pattern similarity between trainees × athletes vs. trainees × matched controls; Fig. 3a, see section "trainee-athlete RSA" in the Methods).

During word list encoding, we observed significantly lower neural pattern similarity in the left superior frontal gyrus when trainees had completed the method of loci training (repeated measures ANOVA, $n = 17$, main effect of *time point*: F (1, 16) = 19.001, $p < 0.001$, $\eta_p^2 = 0.54$, and significantly lower neural pattern similarity between trainees × athletes as compared to between trainees × matched controls (main effect of *group*: F (1, 16) = 29.577, $p < 0.001$, $\eta_p^2 = 0.65$). We also found a significant *time point* × *group* interaction (F (1, 16) = 12.080, $p < 0.001$, $\eta_p^2 = 0.43$). Simple effects showed that neural pattern similarity between trainees × athletes ($p = 0.002$) and between trainees × matched controls ($p = 0.001$) was lower in the post-training compared to the pre-training session (Fig.3b). Whole-brain searchlight analysis confirmed these findings (flexible factorial analysis with the factors *time point* and *group*, $n = 17$, $p < 0.05$, FWE-corrected, cluster-defining threshold $p < 0.001$, critical cluster size = 38 voxels, Fig.3c, Table S4). These effects were only present in the memory training group but not in the active or passive control groups (Fig. S4). Therefore, results suggest that method of loci training leads to distinct neural representations between trainees and memory athletes when studying novel words.

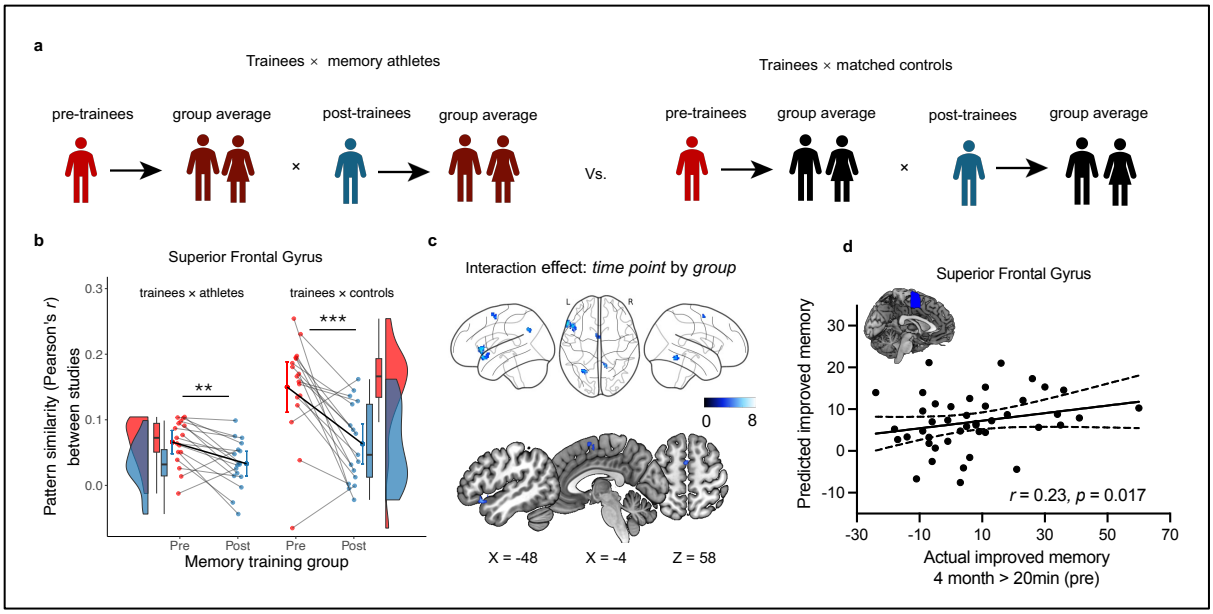

Fig. 3. Method of loci training yields distinct neural representations between trainees and memory athletes during word list encoding. (**a**) Detailed schematic of the trainee-athlete RSA: We calculated the trainee-to-athletes ("trainees" are individuals of the memory training group) and trainee-to-controls neural pattern similarity during word list encoding and compared between pre-(red) and post-training (blue) time points (see Methods). (**b**) ROI results (control analyses for the active and passive control groups are shown in Fig. S4). Data points represent individual participants (memory training group, $n = 17$). Boxplots display the median and upper/lower quartile, with whiskers extending to the most extreme data points within 1.5 interquartile ranges above/below the quartiles. The data points with error bars show the mean ± S.E.M. Data distributions are based on the respective probability density function. (**c**) Searchlight results for the time point × group interaction. Results were thresholded at $p < 0.05$, FWE-corrected at cluster-level, using a cluster-defining threshold of $p < 0.001$ (Table S4). (**d**) The scatter plot depicts the correlation between memory performance at the 4-month reset (4 months > immediate free recall test during the pre-training session) and the predicted changes in neural pattern similarity between all individuals of the training study and memory athletes from pre- to post-training ($n = 44$). ***$p < 0.001$; **$p < 0.01$; ns, not significant.

Distinct neural representations during word list encoding are associated with better free recall performance after 4 months

We then tested the behavioral relevance of the distinct neural representations between all individuals of the training study and memory athletes at the time of encoding with respect to longer-term memory performance. Individuals of the training study were re-invited to the behavioral laboratory to participate in a re-test after 4 months. During this session, they completed the word list encoding task once more, followed by a free recall test, which we leveraged to derive a measure of free recall performance (see Methods). Our behavioral data showed that only the athletes and memory training group exhibited a significant improvement in free recall performance from baseline to the 4-month re-test, while no such improvement was observed in the other groups (see memory performance per group in Table. S8). To further explore the relationship between neural representations and long-term memory improvement, we included all individuals of the training study (i.e., all three groups) in the prediction analysis, rather than restricting it to the method of loci group alone. This inclusive approach ensures that the analysis captures both the specific effects of method of loci training and broader training-related mechanisms across all groups, while enhancing statistical power and generalizability. We next employed a cross-validated machine-learning approach (Cohen, 2010; Supekar et al., 2013) to predict the increase in free recall performance due to method of loci training (i.e., the change in free recall performance from baseline (measured 20 minutes pre-fMRI) to the 4-month re-test) based on the changes in neural pattern similarity between trainees × athletes from pre- to post-training during word list encoding (see section "Prediction analysis" in the Methods and trainee × athlete effects in Fig. 3a).

The results showed that neural pattern similarity between all individuals of the training study and memory athletes in the left superior frontal gyrus was significantly associated with better free recall performance after 4 months ($r_{(\text{predicted, observed})} = 0.23$, $p = 0.017$, $n = 44$, Fig. 3d). We also repeated the analysis by defining the change in free recall performance after 24h, which gave virtually identical results (Table. S9). In summary, method of loci training led to distinct, encoding-related neural representations between experienced individuals that expressed a longer-lasting benefit on memory performance.

Memory training is associated with shared neural representations between experienced individuals during temporal order recognition

Following word list encoding, participants engaged in a temporal order recognition task where they were presented with word triplets from the previously studied word lists and were asked to indicate whether words were presented in the same or in a different order as during study (Fig. 4a). Memory athletes and individuals from the memory training group (post-training) were instructed to employ the method of loci during this task. It is important to highlight that this analysis can yield novel insights as to whether method of loci training affects general retrieval processes during temporal order recognition rather than focusing on the "uniqueness" of each word-location association (as we have done so far). This is because temporal order judgments were based on word triplets (rather than sequences of single words), thereby amplifying the neural processes that are shared across the different items (i.e., the analysis picks up upon common, retrieval-related aspects that are similar across different words). Nevertheless, based on our results from word list encoding, we expected distinct neural representations (thus, lower neural alignment) between memory athletes and between individuals of the memory training group after training. We focused our analysis on brain regions that we previously found to be associated with temporal order recognition following method of loci training, such as the hippocampus and precuneus (Wagner et al., 2021). Between-subject RSA was performed identically to the word list encoding task, quantifying neural pattern similarity by correlating the ROI-based (Fig.4b) neural patterns within each of the experimental groups (e.g., memory athletes × memory athletes, and memory training group × memory training group, separately for pre- and post-training; Fig. 2a, and see section "between-subject RSA" in the Methods).

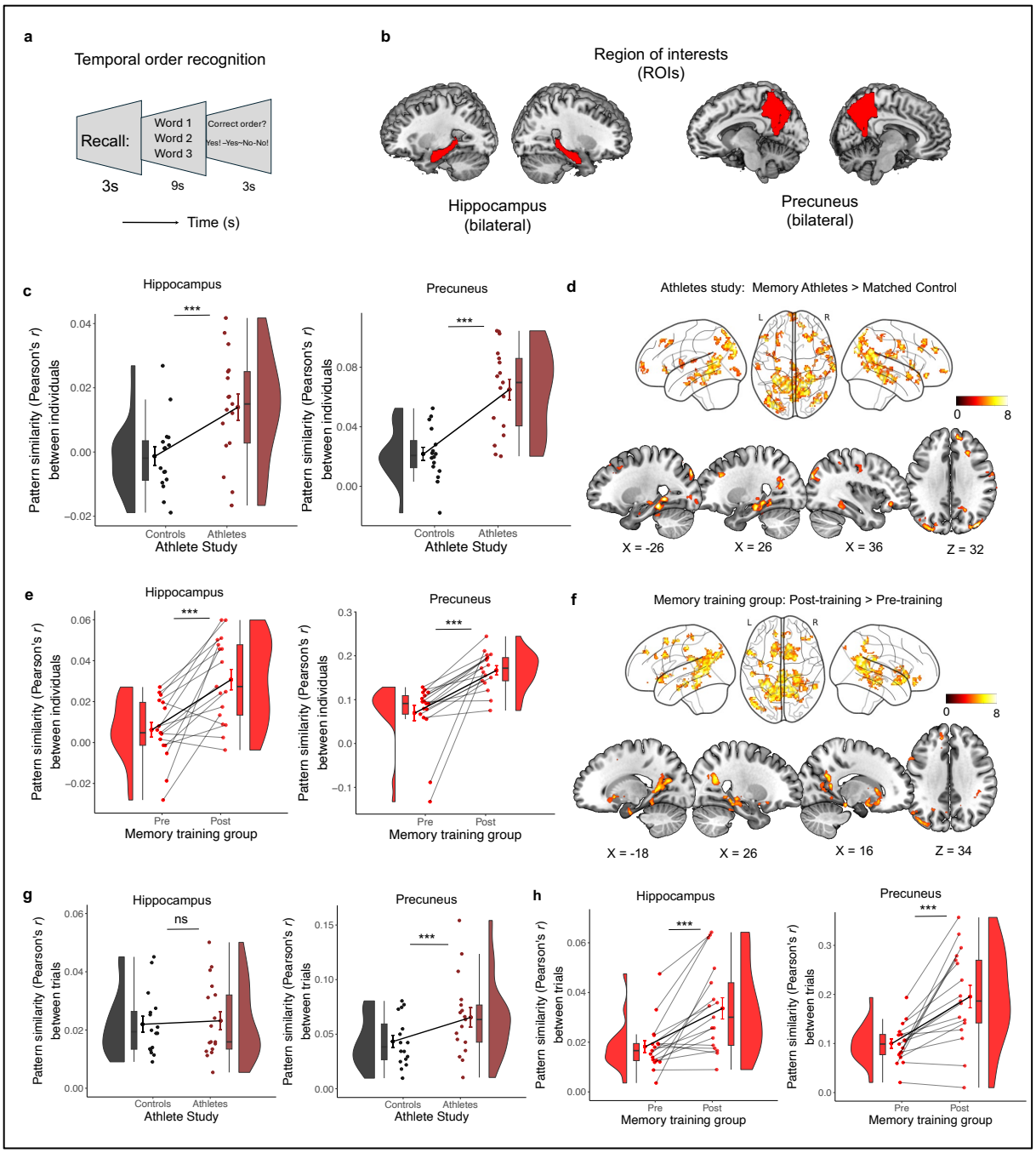

Fig. 4. Method of loci training yields shared neural representations between experienced individuals during temporal order recognition. (**a**) Schematic of the temporal order recognition task. (**b**) ROIs that were used for all analyses pertaining to temporal order recognition. (**c, e**) ROI results of the between-subject RSA (control analyses for the active and passive control groups are shown in Fig. S5). (**d, f**) Searchlight results of the between-subject RSA. Results were thresholded at $p < 0.05$, FWE-corrected at cluster-level, using a cluster-defining threshold of $p < 0.001$ (Tables S5-6). (**g, h**) ROI results of the trial-specific RSA pertaining to the (**g**) athlete study and (**h**) the training study (control analyses for the active and passive control groups are shown in Fig. S7). (**c, e, g, h**) Data points represent individual participants (athlete study, $n = 33$, memory training group, $n = 17$). Boxplots display the median and upper/lower quartile, with whiskers extending to the most extreme data points within 1.5 interquartile ranges above/below the quartiles. The data points with error bars show the mean ± standard error of the mean (S.E.M.). Data distributions are based on the respective probability density function. All ROI-based results survived multiple comparison corrections using the false discovery rate (FDR) with the Benjamini–Hochberg procedure set to $q < 0.05$. ***$p < 0.001$; ns, not significant.

Contrary to what we had expected, results showed significantly increased neural pattern similarity between memory athletes compared to matched controls in the hippocampus (separate independent-sample *t*-tests, memory athlete, $n = 17$, matched controls, $n = 16$, $t(31) = 6.430$, $p < 0.001$, Cohen's $d = 2.240$; Fig.4c) and in the precuneus during temporal order recognition ($t(31) = 4.731$, $p < 0.001$, Cohen's $d = 1.648$; Fig.4c). Whole-brain searchlight analysis showed additional increases in neural pattern similarity in the parahippocampal-, fusiform-, and in the superior temporal gyrus, as well as in the angular gyrus (independent-sample *t*-test, same samples as above, contrast recognition > baseline, memory athletes > matched controls, $p < 0.05$, FWE-corrected, cluster-defining threshold $p < 0.001$, critical cluster size = 35 voxels, Fig. 4d, Table S5).

Similarly, we observed significantly increased neural pattern similarity between individuals of the memory training group after training (compared to the pre-training baseline) in the hippocampus (separate paired-sample *t*-tests, $n = 17$, $t(16) = 4.842$, $p < 0.001$, Cohen's $d = 1.174$; Fig. 4e) and in the precuneus ($t(16) = 4.852$, $p < 0.001$, Cohen's $d = 1.177$; Fig. 4e). This was not the case in the active or passive control groups (Fig. S5). Whole-brain searchlight analysis confirmed these results and further showed increased pattern similarity in the parahippocampal-, and fusiform gyrus, the caudate, and angular gyrus, among other regions (paired-sample *t*-test, $n = 17$, contrast recognition > baseline during post-training > pre-training session, $p < 0.05$, FWE-corrected, cluster-defining threshold $p < 0.001$, critical cluster size = 30 voxels, Fig. 4f, Table S6). These effects were larger in the memory training group compared to both the active and passive control groups (Fig. S6). In summary, we found shared neural representations between memory athletes and between individuals of the memory training group (post-training) as they recognized the temporal order of previously studied material.

Experienced individuals hold similar neural representations associated with different word triplets during temporal order recognition

Once again, we tested whether the finding of shared neural representations between experienced individuals was perhaps attributable to each individual holding similar neural representations associated with the different word triplet during temporal order recognition. We performed trial-specific RSA identical to the analysis of the word list encoding task, which yielded the average neural pattern similarity across all word triplets per participant and for each *a priori* defined ROI (see section "Trial-specific RSA" in the Methods).

Findings showed significantly increased neural pattern similarity across word triplets in memory athletes compared to matched controls in the precuneus (separate independent-sample *t*-tests, memory athlete, $n = 17$, matched controls, $n = 16$, $t(31) = 4.026$, $p < 0.001$, Cohen's $d = 1.402$; Fig. 4g), but not in the hippocampus during temporal order recognition ($t(31) = 0.568$, $p = 0.574$, Cohen's $d = 0.198$; Fig. 4g). In a similar vein, data from the training study showed significantly increased neural pattern similarity across word triplets in the memory training group after training (compared to the pre-training baseline) in both the hippocampus (separate paired-sample *t*-tests, $n = 17$, $t(16) = 4.372$, $p < 0.001$, Cohen's $d = 1.060$; Fig. 4h) and precuneus ($t(16) = 4.890$, $p < 0.001$, Cohen's $d = 1.186$; Fig. 4h). This was not the case in the active or passive control groups (Fig. S7). Thus, the shared neural representations between individuals who were proficient in the method of loci appeared driven by similar neural representations of the different word triplets during temporal order recognition.

Shared neural representations between trainees and memory athletes during temporal order recognition

Thus far, we showed shared neural representations when individuals of the memory training group after training (the so-called "trainees"), or memory athletes, were judging the temporal order of previously studied words. We expected that these effects would translate to the between-group level, showing shared neural representations between trainees and memory athletes. This would indicate that the retrieval-related processes engaged in correctly judging the temporal order of the unique word-location associations are comparable between individuals who are proficient in the method of loci. To quantify the degree of trainee-

athlete similarity in neural representations, we performed the abovementioned focused on the respective ROIs (Fig. 3a, see section "trainee-athlete RSA" in the Methods)

Results showed significantly increased neural pattern similarity in the precuneus when trainees had completed the method of loci training (repeated measures ANOVA, $n = 17$, main effect of *time point*: $F(1, 16) = 6.016$, $p = 0.026$, $\eta_p^2 = 0.27$), and significantly higher neural pattern similarity between trainees × athletes compared to pattern similarity between trainees × matched controls (main effect of *group*: $F(1, 16) = 9.775$, $p = 0.007$, $\eta_p^2 = 0.38$). We also found a significant *time point* × *group* interaction ($F(1, 16) = 92.430$, $p < 0.001$, $\eta_p^2 = 0.85$), showing that neural pattern similarity between trainees × athletes ($p = 0.002$) but not between trainees × matched controls ($p = 0.208$) was higher in the post- compared to pre-training session (Fig.5a; results remained unchanged when removing two formal outliers in the pre-training session). Whole-brain searchlight analysis confirmed these findings (flexible factorial analysis with the factors *time point* and *group*, $n = 17$, $p < 0.05$, FWE-corrected, cluster-defining threshold $p < 0.001$, critical cluster size = 37 voxels, Fig.5b, Table S7). Again, these effects were only present in the memory training group but not in the active or passive control groups (Fig. S8). Together, these results show that method of loci training leads to shared neural representations between experienced individuals (i.e., trainees and memory athlete) during temporal order recognition.

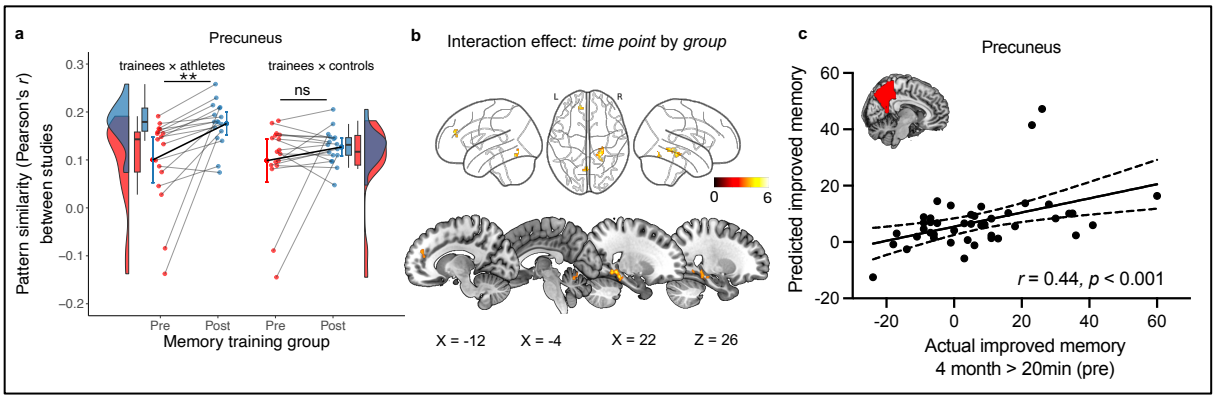

Fig. 5. Method of loci training yields shared neural representations between trainees and memory athletes during temporal order recognition. (**a**) ROI results of the trainee-athlete RSA (control analyses for the active and passive control groups are shown in Fig. S8). Data points represent individual participants (memory training group, $n = 17$). Boxplots display the median and upper/lower quartile, with whiskers extending to the most extreme data points within 1.5 interquartile ranges above/below the quartiles. The data points with error bars show the mean ± S.E.M. Data distributions are based on the respective probability density function. (**b**) Searchlight results for the time point × group interaction. Results were thresholded at $p < 0.05$, FWE-corrected at cluster-level, using a cluster-defining threshold of $p < 0.001$ (Table S7). (**c**). The scatter plot depicts the correlation between memory performance at the 4-months reset (4 months > immediate free recall test during the pre-training session) and the predicted changes in neural pattern similarity between all individuals of the training study and athletes from post- to pre-training ($n = 44$, see Methods). **$p < 0.01$; ns, not significant.

Shared neural representations during temporal order recognition are associated with better free recall performance after 4 months

In a final step, we tested the behavioral relevance of the shared neural representations between experienced individuals during temporal order recognition with respect to longer-term memory performance. As above, we employed a cross-validated machine-learning approach (Cohen, 2010; Supekar et al., 2013) to predict the increase in free recall performance due to method of loci training (i.e., the change in free recall performance from baseline to the 4-month re-test) based on the changes in neural representations between all individuals of the training study × athletes from pre- to post-training during temporal order recognition (see section "Prediction analysis" in the Methods and trainee × athlete effects in Fig. 3a). Results showed that increased neural pattern similarity between all individuals of the training study and athletes in the

precuneus was significantly associated with better free recall performance after 4 months ($r_{(\text{predicted, observed})}$ = 0.44, $p < 0.001$, $n = 44$, Fig. 5c). We also repeated the analysis by defining the change in free recall performance after 24h, which gave virtually identical results (Table.S9). In summary, method of loci training led to shared, retrieval-related neural representations between experienced individuals that predicted a longer-lasting benefit on memory performance.

Discussion

We investigated the impact of a 6-week-long method of loci training on the neural representations that support superior memory. To characterize whether the effects of such multiple-week training would be comparable to existing expertise with mnemonic techniques, we performed two separate studies that involved initially mnemonics-naïve individuals who completed a method of loci training, and experienced memory athletes. Several key results emerged from the data that substantially expand our previous work (Dresler et al., 2017; Wagner et al., 2021): We found that method of loci training is associated with distinct neural representations between experienced individuals as they studied lists of novel words. Put differently, the memory training group (post-training) as well as memory athletes appeared less similar in their encoding-related neural representations. This was also reflected at the level of the studied material, showing that individuals formed distinct neural representations associated with each word-location association. Focusing on the trainee-athlete similarity during encoding, we showed that more distinct neural representations exerted a lasting benefit on memory as they predicted free recall performance after 4 months. This may be related to the formation of unique word-location associations in an individual's "memory palace". We also took the opportunity to explore the neural representations during temporal order recognition. There, memory training led to shared (rather than distinct) neural representations between experienced individuals, suggesting an impact on associative retrieval processes. Taken together, our findings provide first insights into how extended memory training transforms the neural representations that enable effective, superior memory.

Our focus revolved around the potential impact of method of loci training on the encoding-related neural representations in individuals of the memory training group after they completed their training, as well as in memory athletes. We found distinct neural representations between these experienced individuals in the left lateral prefrontal cortex and superior frontal gyrus during word list encoding. This was mirrored at the level of the studied material, revealing that experienced individuals held unique neural representations that were associated with the specific words. In line with these results, Liu and colleagues (2022) reported that 5-day method of loci training was associated with lower neural pattern similarity in hippocampal subfields during memory encoding (thus, distinct hippocampal patterns). This effect was specific to the studied material that was mentally placed at close-by locations along the imagined route and was coupled to accurate temporal order memory. Bellmund and colleagues (2022) investigated neural pattern similarity during the encoding of event sequences (but their study did not involve memory training). The authors observed distinct hippocampal patterns when events appeared closer in time. More evidence comes from LaRocque and co-authors (2013) who reported that successful recognition memory was predicted by distinct hippocampal representations at encoding. Here, we found that memory training led to distinct neocortical representations in lateral and superior prefrontal regions. The lateral prefrontal cortex is known as a key player when it comes to reducing memory interference during learning (Ghazizadeh et al., 2018; Sakai et al., 2002). Using electrophysiological recordings in non-human primates, Ghazizadeh and colleagues (2018) showed that lateral prefrontal neurons coded for object-reward associations that persisted over several months, that were resistant to extinction, and that were not affected by new learning. Adjacent inferior prefrontal regions were shown to support the retrieval of detailed episodic information (Badre & Wagner, 2007; Öztekin et al., 2009) and to differentiate (temporal) contexts in both non-human primates (Mante et al., 2013) and humans (Sakai et al., 2002; Schlichting et al., 2015). This resonates with theoretical accounts of pattern separation that highlight the role of distinct neural representations in orthogonalizing overlapping memory content and minimizing interference (Amer & Davachi, 2023; Nash et al., 2021; Stevenson et al., 2020; Yassa & Stark, 2011), as well as with the role of prefrontal regions in supporting

437 the longer-term stabilization of memories via top-down projections (Blumenfeld & Ranganath, 2007).
438 Together, our findings suggest that method of loci training may amplify memory "uniqueness" during
439 encoding by reducing the spatiotemporal similarity between events, fostering memory persistence.

440 Continuing along these lines, the formation of maximally unique associations appears to be a key aspect of
441 successful memory formation when using the method of loci. Mentally placing to-be-remembered
442 information at salient locations along the imagined path likely produces relatively bizarre associations,
443 thereby triggering neural mechanisms related to novelty (Fernández & Morris, 2018). This is known to
444 crank up dopamine and noradrenaline release from the brainstem and ventral striatum, which boosts
445 memory persistence (Duszkiewicz et al., 2019). In addition, the lateral prefrontal cortex appears associated
446 with the processing of novel information as well (Ranganath & Rainer, 2003). Prior work using invasive
447 recordings in non-human primates showed that novel stimuli increase the activity of neurons in the lateral
448 prefrontal cortex (Matsumoto et al., 2007). More evidence comes from lesion studies, highlighting that
449 damage to the lateral prefrontal cortex abolishes the memory advantage for novel compared to non-novel
450 stimuli in both non-human primates (Parker et al., 1998) and human patients (Kishiyama et al., 2009). Here,
451 we found distinct neural representations in the lateral prefrontal cortex between experienced individuals
452 during encoding, which might be related to the processing of novel word-location associations. Our results
453 further included the superior frontal gyrus. Human patients with lesions in this area exhibit marked deficits
454 in working memory and spatially oriented attention (Boisgueheneuc et al., 2006). Other reports discussed
455 this region's role in monitoring and executive processing (Owen, 2000; Petrides, 2000), as well as its part
456 within a prefrontal hierarchy that processes multi-feature and stimulus-response associations along an
457 inferior-to-superior gradient (Koechlin et al., 2003). We thus speculate that the prefrontal involvement
458 during word list encoding might stem from the formation of complex, multi-feature associations (lateral
459 prefrontal cortex) that require monitoring and spatial attention (superior frontal gyrus). Since the memory
460 training group and memory athletes received extensive training in utilizing the method of loci, the process
461 of associative memory formation might become comparable to stimulus-response learning (Müller et al.,
462 2018), allowing for the rapid integration of novel information into the well-established "memory palace".

463 We also demonstrated distinct neural representations between experienced individuals in posterior
464 representational regions during encoding, including the left inferior temporal cortex and bilateral posterior
465 parietal areas. The inferior temporal cortex was shown to code for object categories during perception (Op
466 de Beeck et al., 2008) and mental imagery (Dijkstra et al., 2019). Previous work that employed invasive
467 recordings in monkeys and fMRI-based recordings in humans demonstrated that the representational
468 geometry in the inferior temporal cortex (e.g., how distinct the neural representations of two objects are)
469 depends on an object's visual features (Kriegeskorte, Mur, Ruff, et al., 2008) and is highly specific to an
470 individual's perceptual judgments (Charest et al., 2014). Moreover, our results extend toward the (left)
471 visual word form area and to posterior parietal regions, which are known to selectively process written letter
472 strings (Dehaene & Cohen, 2011) and to code for (egocentric) spatial context (Baddeley et al., 2001; Sereno
473 et al., 2001; Takashima et al., 2007), respectively. The visual presentation of the words during word list
474 encoding may trigger vivid imagery of unique word-location associations, associated with distinct neural
475 representations that may orthogonalize memory content (Yassa & Stark, 2011). Naturally, these effects are
476 likely specific to the stimulus material at hand and might be different for other sensory modalities and task
477 setups.

478 Our findings of distinct neural representations between experienced individuals stand in contrast to previous
479 reports that highlighted the beneficial effect of interindividual similarity on cognitive performance. For
480 instance, Koch and colleagues (2020) demonstrated that successful scene encoding was coupled to
481 increased neural pattern similarity between individuals. Sheng and colleagues (2023) showed that the
482 degree of neural pattern similarity to an average, group-based neural pattern during associative encoding
483 predicted subsequent memory ability across individuals. Said differently, the more individuals resembled
484 the group in their neural representations, the better their memory was. Such "neural alignment" might stem
485 from the processing of similar content. Therefore, it follows logically that the processing of unique memory

content (as is the case when utilizing the method of loci) is tied to distinct neural representations. We next were curious to test whether the distinct neural representations between experienced individuals would also translate to the relationship between trainees (i.e., individuals of the memory training group after training) and memory athletes. In their previous study, Meshulam and colleagues (2021) not only showed neural alignment between learners after acquiring new conceptual knowledge but also reported neural alignment between learners and experts that was predictive of learning outcomes. The authors interpreted their effects in terms of group knowledge, or an ideal canonical representation, comparable to memories for shared experiences (Chen et al., 2017b). In line with our results of distinct neural representations between experienced individuals, we also found distinct neural representations between trainees and memory athletes. Trainees and memory athletes were less similar in their encoding-related neural representations than individuals who did not undergo memory training. While the acquisition of a shared representation might certainly be useful when conceptual knowledge structures are acquired, we suggest that distinct neural representations may be favorable when it comes to memory specificity. For instance, unique memories were shown to be retrieved with increased visual detail compared to repeated events or facts (Tanguay et al., 2023). The distinct trainee-athlete neural representations also predicted free recall performance after 4 months, emphasizing the beneficial impact of memory "uniqueness" on memory persistence (Burgess et al., 2002; Ranganath & Ritchey, 2012).

Following the word list encoding task, participants completed a temporal order recognition task. This allowed us to explore whether method of loci training had comparable effects on neural representations when individuals retrieved the previously learned word-location associations. It is important to note that temporal order judgments were based on word triplets (rather than sequences of single words). Hence, the analysis is not directly comparable with the encoding-based results as it amplifies neural processes that are shared across the different items (i.e., the analysis picks up upon common, retrieval-related aspects that are similar across different words). Indeed, results showed that memory training led to shared (rather than distinct) neural representations between experienced individuals in the hippocampus and precuneus, as well as in the posterior parahippocampal gyrus, retrosplenial cortex, and angular gyrus. Temporal order recognition likely involves the reinstatement of neural representations from prior encoding. Episodic memory reinstatement was associated with fMRI-based neural representations in the hippocampus, posterior midline, and parietal regions (Oedekoven et al., 2017), as well as with direct hippocampal recordings in human epilepsy patients (Pacheco Estefan et al., 2019). Additionally, the hippocampus is involved in comparing incoming information with prior knowledge, as described by theoretical frameworks of pattern completion (Hindy et al., 2016; Horner et al., 2015). The precuneus is involved in integrating self-referential (Lyu et al., 2023) and episodic memory processes (Sestieri et al., 2017) and is engaged in recalling temporal sequences. Xiao and co-authors (2020) reported shared neural representations between individuals during associate retrieval in the angular gyrus. We propose that the shared neural representations between individuals reflect common retrieval processes, such as content reinstatement and pattern completion, that are similar across word triplets.

In conclusion, we provide the first evidence of how multiple-week memory training transforms the neural signatures that underlie superior memory. We found distinct neural representations between memory athletes and between individuals who were trained to utilize the method of loci over several weeks (the so-called "trainees") while they studied novel material. The distinct neural representations were also evident between trainees and memory athletes, and more distinct representations predicted better memory performance after 4 months. We suggest that these effects are linked to pattern separation and novelty processes that bolster individual memory "uniqueness". We also investigated neural representations during temporal order recognition, highlighting the impact of memory training on general retrieval-related processes in the form of shared neural representations between experienced individuals. Collectively, we suggest that method of loci training serves to disambiguate memory content upon encoding, creating unique memories within one's "memory palace" that subserve extraordinary memory ability.

Acknowledgments

We thank S. Weisig and P. Schuster for their help with data collection and the International Association of Memory (IAM; www.iam-memory.org) for support in recruiting memory athletes. This article is based on work from a Veni project supported by the Dutch Research Council (NWO); and the European Platform for Life Sciences, Mind Sciences, and the Humanities supported by the Volkswagen Foundation. I.C.W. is supported by the Austrian Science Fund (FWF, 10.55776/P34775) and by the European Research Council (Starting Grant 101164099).

Author contributions

M.D., B.N.K. and M.C. proposed the original study design. B.N.K. and M.D. acquired the original data. J.R., I.C.W. and M.D. conceived the specific research question at hand. J.R. and I.C.W. contributed to methods development. J.R. and I.C.W. analyzed the data. J.R., I.C.W., Y.Z., F.L. and M.D. contributed to the interpretation of the results. J.R. and I.C.W. drafted the paper. All authors contributed to the critical revision of the paper.

Competing interests

The authors declare that they have no competing interests.

Methods

Participants of the athlete and training studies

In the athlete study, we enrolled 23 memory athletes (mean age ± s.d., 28 ± 8.6 years, 9 females) who were among the top 50 participants in the World Memory Championships at the time of study participation (www.world-memory-statistics.com). Of these 23 athletes, 17 took part in a word list encoding and temporal order recognition task inside the MRI scanner. Memory athletes were matched to a control group that was selected from gifted students of academic foundations and members of the high-IQ society "Mensa" (https://www.mensa.org/). Matching criteria included IQ, age, sex, handedness, and smoking status. Six participants of the matched control group were selected from the training study based on their cognitive abilities, as determined during the screening session (details of the screening session are provided below). They were evenly sampled from the three training groups. All participants of the matched control group underwent a standardized memory test (Bäumler, 1974) to prevent the inclusion of "natural" superior memorizers (none of the participants met this criterion). Additionally, they completed a test for fluid reasoning (Weiß, 2006). Any prior experience with systematic memory training was an exclusion criterion. Together, these participant groups were part of the so-called "athlete study" (memory athletes: $n = 17$, mean age ± s.d., 24.6 ± 4.3 years, 8 females; matched controls: one participant was excluded due to excessive head motion during fMRI, $n = 16$, mean age ± s.d., 25.4 ± 3.9 years, 7 females).

To assess the behavioral and neural effects of mnemonic training in participants without prior experience with the method of loci, we enrolled 51 healthy, right-handed individuals (mean age ± s.d., 24 ± 3 years) from the University of Munich. This constituted the training study. We only included male participants due to the known influence of the menstrual cycle on memory (Genzel et al., 2012; Sattari et al., 2017) and hippocampal morphology (Zsido et al., 2023) and because the longitudinal study design would not have allowed for a systematic control of this factor. None of the participants had any reported experience with memory training, and none exhibited superior memory abilities. Participants were assigned to one of the three groups based on their cognitive performance as determined during the initial screening session (using assessments from Bäumler, 1974; Weiß, 2006). This allowed us to ensure that any potential changes in memory performance would be attributable to the method of loci training (see Table S1). The memory training group underwent a 6-week training in the method of loci ($n = 17$, mean age ± s.d., 24 ± 3 years). These participants were directly compared to a group that completed a working memory training (active controls: one participant was excluded because of technical issues with the MR recordings, $n = 16$, mean age ± s.d., 24 ± 3 years) and to a group that did not undergo any intervention (passive controls: n = 17, mean age ± s.d., 24 ± 4 years).

All participants were native German speakers with normal or corrected-to-normal vision, and without a history of psychiatric or neurological diseases. Written informed consent was obtained before study enrollment, and the study protocol was reviewed and approved by the ethics committee of the Medical Faculty of the University of Munich, Germany.

Study overview and cognitive training

Participants of the athlete study completed a single MRI session (Fig. 1a). Participants of the training study completed two identical MRI sessions that were separated by a 6-week interval, as well as a behavioral re-test four months later (Fig. 1b).

Method of loci training

The memory training group completed systematic memory training using the method of loci across 6 weeks. Participants were introduced to the method of loci at the Max Planck Institute of Psychiatry (Munich, Germany). They were guided in constructing their first loci route both within and outside the institute and received supervision while using the route in an initial memory task. Subsequently, participants were

acquainted with the online platform that was utilized for home-based training, for monitoring training progress, and compliance (https://memocamp.com). They were also provided with a training plan for the upcoming week.

After that, participants underwent a 40-day home-based training (30 minutes per day). The training plan included instructions for selecting locations to reduce interference from previously memorized word lists and to ensure an equal distribution of training time across different routes. During the first two weeks of training, participants were instructed to create and memorize three additional loci routes, resulting in a total of four routes, including the one introduced during the orientation session. These routes were employed for memorizing random words. The task difficulty, which was determined by the number of words to be memorized, was adjusted dynamically based on individual performance. For example, at the beginning of each daily training session, participants were presented with five words during the first run. If participants successfully recalled all the words in a given run, the number of presented words increased by +5 in subsequent runs. The speed of training success was defined as the average number of runs needed to progress to the next difficulty level, and 40 words needed to be recalled successfully (equivalent to eight runs). While most participants in the memory training group reached this final level (16 out of 17), it is hardly achievable for mnemonics-naïve participants.

Log files from the daily training sessions were monitored daily to ensure participants' compliance. In cases where a participant missed a training session or did not complete their daily routine, they were contacted the following morning. They then received instructions to increase their training time on the subsequent day to compensate for the missed session. Additionally, participants were required to visit the laboratory once a week for supervised training in small groups of 2-3 participants. During these sessions, they were interviewed about their training regime and received a training plan for the upcoming week.

Working memory training and passive controls

Participants of the active control group were introduced to the dual *n*-back task that required them to monitor and update a series of visually presented locations and auditorily presented letters (Jaeggi et al., 2008). The value of n changed between sets of trials, continuously adjusting based on performance. This adaptive process resulted in the task demand changing according to each participant's individual performance. They engaged in a 40-day home-based training (30 minutes per day, using an online working memory training program). To ensure compliance, their progress was monitored daily. If participants missed or did not complete their daily training routine, they were notified the following morning, with instructions to increase their training time on the subsequent day to compensate for the missed session. Additionally, participants visited the laboratory once a week in small groups of 2-3 participants for an interview regarding any potential training issues and received a training under direct supervision. Participants were instructed to perform to the best of their ability but were advised to avoid any systematic long-term memory training. The passive control group did not receive any training between the two sessions and were given no specific instructions beyond refraining from systematic memory training while being enrolled in the study.

General structure of MRI sessions

An MRI session of both the athlete and training studies, during pre- and post-training began with the acquisition of a structural brain image, followed by a (baseline) resting-state fMRI period, the word list encoding and temporal order recognition tasks (as shown in Fig. 1c), and a post-task resting-state fMRI (resting-state periods are not further discussed in this paper). Subsequently, participants completed a free recall test in the behavioral laboratory 20 minutes after leaving the MR scanner (i.e., the immediate free recall test), and another free recall test via a phone interview 24 hours later (i.e., the delayed free recall test). Participants of the athlete study only performed the immediate free recall test but not the delayed free recall test.

640 ***Word list encoding task***
641 We utilized this task (along with the temporal order recognition task described below), because we reasoned
642 that the particular strength of the method of loci lies in learning and recalling ordered sequences, facilitated
643 by the mental navigation through an imagined "memory palace". The task involved concrete nouns that
644 were divided into two separate lists (72 words each), that were counterbalanced for word length and
645 frequency and were presented in random order (pre- and post-training, and the order of lists was also
646 balanced across participants).

647 The task began with a brief instruction (5 seconds), after which the first word was presented (3 seconds),
648 followed by a fixation cross displayed for a variable duration ranging from 2-5 seconds (mean = 3.5
649 seconds). A 30-second fixation period was inserted after every sixth word. Participants were explicitly
650 instructed not to rehearse the studied material during these fixation periods but to keep their thoughts
651 unoccupied. The total duration of the task was 10 minutes. All stimulus material was presented with
652 Presentation (Neurobehavioral Systems, Albany, CA, USA). Memory athletes and the memory training
653 group (post-training) were instructed to employ the method of loci to memorize the word lists during the
654 task. In other words, participants were asked to mentally navigate through their "memory palace" and to
655 place the different words at their specific loci. The other groups received no explicit instructions.

656 ***Temporal order recognition task***
657 We developed this task to form an MR-compatible measure of recall performance. Participants viewed 24
658 triplets of words based on material from the previously encoded word list (using adjacent words that were
659 shuffled within triplets). Hence, each word triplet consisted of three words that were previously presented
660 in direct (0-distance) and close (1-distance) proximity.

661 A trial commenced with a cue signaling the upcoming word presentation (3 seconds), followed by the
662 presentation of a word triplet (9 seconds). Subsequently, participants were asked to determine whether the
663 word order in the triplet was the same as during the previously encoded word list. They indicated their
664 choice with a button press (answer options "same, sure", "same, maybe", "different, maybe", and "different,
665 sure"; 3 seconds). Triplet presentations were interleaved with a control task during which participants were
666 asked to determine whether novel word triplets were displayed in ascending or descending order based on
667 the number of syllables in each word (using the same timing and response options as described above). The
668 word order recognition trials (A) and the control task (B) were presented in an ABAB fashion. The total
669 duration of the task was 10 minutes. Memory athletes and the memory training group (post-training) were
670 instructed to use the method of loci to judge whether the presented triplets were presented in the correct
671 order. In other words, participants were asked to mentally navigate through their "memory palace",
672 retrieving individual words to assess whether they were presented in the correct order. No specific
673 instructions were given to the other groups.

674 ***Free recall***
675 The free recall test served as our primary outcome measure of memory performance, as it closely resembles
676 tasks used in memory championships. Following MR scanning (approximately 20 minutes later),
677 participants were instructed to freely recall and write down the 72 concrete words they had learned during
678 the preceding word list encoding task (immediate free recall test). After 5 minutes, participants were asked
679 if they needed more time, and if so, the free recall was terminated after an additional 5 minutes. The delayed
680 free recall test was conducted 24 hours later via phone, following the same 5+5-minute format. Memory
681 performance was assessed by counting the number of words correctly recalled, ignoring word order and
682 spelling mistakes. The delayed free recall test was announced to all participants, as our intention was to
683 maintain consistency between pre- and post-training sessions. Notably, participants in the athlete study
684 completed only the immediate free recall test and not the delayed version.

685 ***Re-test after four months***
686 Participants in the training study also completed a behavioral re-test four months after the post-training
687 session. During this re-test, we performed the word list encoding task, followed by a delay period that

included a reasoning task (15 minutes) and a free recall task (5+5 minutes, serving as our primary outcome measure). The task material during the word list encoding task was identical to that used during the initial pre-training session because the 4-month re-test was added after the word lists had been constructed. Importantly, participants were assigned to the different training groups only after the pre-training session. As a result, any material from the first session that might have been remembered during the 4-month re-test should be unaffected by training. As mentioned earlier, participants in the memory training group (post-training) were instructed to use the method of loci during word list encoding. Five participants (2 from the memory training group, 2 from the active control group, and 1 from the passive control group) were not available for the re-test.

Analysis of free recall performance

To quantify memory performance, we calculated the change in freely recalled words from pre- to post-training sessions (post- minus pre-training). This was done separately for the immediate free recall test (20 min post- minus pre-training), the delayed free recall test (24 hours post- minus pre-training), as well as for the re-test after four months (4-month re-test minus pre-training test 20 min). Memory athletes only completed the immediate free recall test (20 min post-MRI) and their performance was quantified as the number of recalled words.

MRI data acquisition parameters

All imaging data were collected at the Max Planck Institute of Psychiatry (Munich, Germany) using a 3T scanner (GE Discovery MR750, General Electric, USA) with a 12-channel head coil. We obtained a time series of 292 T2*-weighted BOLD images with the following EPI sequence: TR, 2.5 s; TE, 30 ms; flip angle, 90°; 42 ascending axial slices; FOV, 240 × 240 mm; 64 × 64 matrix; slice thickness, 2 mm, slice gap, 0.5 mm. The structural image was acquired with the following parameters: TR, 7.1 ms; TE, 2.2 ms; flip angle, 12°; in-plane FOV, 240 mm; 320 × 320 × 128 matrix; slice thickness, 1.3 mm.

MRI data preprocessing

The MRI data were processed using Statistical Parametric Mapping (SPM version 12; www.fil.ion.ucl.ac.uk/spm/) in combination with Matlab (The Mathworks, Natick, MA, USA). The first four functional volumes were discarded to allow for T1 equilibrium. The remaining images were corrected for acquisition delay and were realigned to the mean functional image of each session. Processing was done separately for the word list encoding and temporal order recognition tasks. The high-resolution structural image was first segmented into gray matter, white matter, and cerebrospinal fluid, generating a deformation field for spatial normalization. The functional scans were then co-registered to the structural image to ensure alignment. This deformation field, derived from T1 segmentation, was subsequently applied to normalize both the structural and functional images to the standard stereotactic Montreal Neurological Institute (MNI) space. All images were resampled to a 2-mm isotropic voxel resolution.

Modeling of fMRI data

Conventional modeling across trials

To track the alignment of neural representations within and between groups, we employed two separate General Linear Models (GLMs) that captured BOLD activity changes during the word list encoding and temporal order recognition tasks. Using data from the word list encoding task, the BOLD response to all trials was modeled with a single task regressor, time-locked to the onset of each trial (boxcar function with a duration of 3 seconds). Instructions were collapsed into a regressor-of-no-interest (boxcar function with a duration of 5 seconds). Regressors were convolved with the canonical hemodynamic response function (HRF) and the six realignment parameters were included in the model to account for rigid-body head motion. This resulted in a single beta image (capturing average BOLD activity changes during word list

encoding) per participant and session. Using data from the temporal order recognition task, the BOLD response to all trials was modeled as a single task regressor, time-locked to the onset of each trial (boxcar function, duration modeled until a button press occurred). Instructions were incorporated into a regressor-of-no-interest (boxcar function with a duration of 3 seconds). The remaining steps were carried out identically to above. This resulted in a single beta image (capturing average BOLD activity changes during temporal order recognition) per participant and session. For participants of the training study, separate GLM models were defined for the pre- and post-training sessions.

Trial-specific modeling

To track the alignment of neural representations across single words or word triplets, we employed two separate GLMs that captured trial-specific BOLD activity changes during the word list encoding and temporal order recognition tasks. That is, using data from the word list encoding task each presented word was modeled as a separate regressor (following Mumford et al., 2014), time-locked to the onset of each word (boxcar function with a duration of 3 seconds). Instructions were collapsed into a regressor-of-no-interest (boxcar function with a duration of 5 seconds). All regressors were convolved with the canonical HRF and the six realignment parameters were included to account for head motion. This resulted in 72 trial-specific beta images (capturing trial-specific BOLD activity changes during word list encoding) per participant and session. To track the alignment of neural representations across word triplets during the temporal order recognition task, each presented word triplet was modeled as a separate regressor, time-locked to trial onset (boxcar function, duration modeled until a button press occurred). Instructions were grouped into a separate regressor-of-no-interest (boxcar function with a duration of 3 seconds, all remaining details identical to above). This resulted in 24 beta images (capturing trial-specific BOLD activity changes during temporal order recognition) per participant and session. As mentioned above, pre- and post-training sessions were modeled separately for participants of the training study.

Definition of regions-of-interest (ROIs)

We selected four regions-of-interest (ROIs) based on our previous findings of training-related changes in neural activity (Wagner et al., 2021). All ROIs were anatomically defined using the Brainnetome Atlas (https://atlas.brainnetome.org/). For the investigation of changes in neural representations during the word list encoding task, we selected the left superior frontal gyrus (anatomical, left, A6m, medial area 6) and the left lateral prefrontal cortex (anatomical, left, inferior frontal gyrus). For the investigation of changes in neural representations during the temporal order recognition task, we focused on the bilateral hippocampus and precuneus. We included all voxels within the anatomical mask.

Representational Similarity Analysis (RSA)

We employed representational similarity analysis (RSA) to assess the impact of method of loci training on neural representations (Fig.1d). Specifically, we examined (i) the degree of neural alignment between individuals of the same experimental group (i.e., between-subject RSA), (ii) across the studied material (i.e., trial-specific RSA), and (iii) the degree of neural alignment between trainees and memory athletes (i.e., trainee-athlete RSA).

Between-subject RSA: degree of neural alignment between individuals of the same group

This analysis was based on the beta-maps obtained from conventional across-trial modeling (thus, capturing average BOLD activity changes during word list encoding and temporal order recognition). For each participant, neural activity was extracted from the pre-defined ROIs and the data were reshaped into a pattern vector for each ROI. The data were then z-scored across trials to remove mean activation differences. The pattern vectors of each ROI and each participant were then correlated to the pattern vectors of all other participants of the same group, using pairwise Pearson's correlation (r). The resulting correlations were Fisher's z-transformed (using [$z = 0.5*\log(1+r)/(1-r)$]). Between-subject neural pattern similarity for each subject was then computed as the average pattern similarity of each participant to all other participants

within the same group (that is, the average correlation of voxel patterns from participant 1 with those of participants 2~*N* of a given group, Fig 2a). This procedure was repeated for both the pre- and post-training sessions, and separately for the word list encoding and temporal order recognition tasks, as well as separately for the different groups in the athlete and training studies. The following equations were used to quantify the between-subject pattern similarity of individuals in the different experimental groups of each study. In brief, we calculated the pattern similarity of each participant (here denoted as sub_1; to provide an example, we chose a participant of the memory training group, LOC) to the group average of that same group during the pre- (Equation 1) and post-training (Equation 2) sessions during, for example, word list encoding. Equations 3 and 4 follow the same rationale to assess the pattern similarity of a different sub_1 in the matched control (CON) and the memory athlete groups (ATH) to their respective group averages. In these equations, n denotes the total number of participants in the respective groups, and j represents the participant index for participants in athletes or matched control groups other than subject 1, starting from 2 to ensure the exclusion of self-comparison (For details on statistical analysis, see the section "Statistical analyses and thresholding").

$$S_{Pre}(Sub_1, LOC) = \frac{1}{n-1}\sum_{j=2}^{n} r_{Pre}\left(Sub_1, LOC_j\right) \quad (1)$$

$$S_{Post}(Sub_1, LOC) = \frac{1}{n-1}\sum_{j=2}^{n} r_{Post}\left(Sub_1, LOC_j\right) \quad (2)$$

$$S_{ATH}(Sub_1, ATH) = \frac{1}{n-1}\sum_{j=2}^{n} r_{ATH}\left(Sub_1, ATH_j\right) \quad (3)$$

$$S_{CON}(Sub_1, CON) = \frac{1}{n-1}\sum_{j=2}^{n} r_{CON}\left(Sub_1, CON_j\right) \quad (4)$$

Trial-specific RSA: degree of neural alignment across studied material

To investigate whether method of loci training was associated with changes in the neural representation of the studied material, we repeated the abovementioned between-subject RSA procedure but instead based the analysis on the beta-maps that were obtained from the trial-specific modeling (thus, capturing trial-specific BOLD activity changes during word list encoding and temporal order recognition that pertained to the single words and word triplets, respectively). Again, voxel activity patterns were extracted for the pre-defined ROIs for all trials, and data were then *z*-scored across trials to remove mean activation differences. The pattern vectors of each ROI and each trial were then correlated to the pattern vectors of all other trials of the same participant, using pairwise Pearson's correlation (*r*). The resulting correlations were Fisher's *z*-transformed. Trial-specific neural representation for each subject was computed as the average pattern similarities of each trial with all other trials (see Fig. 2h). This procedure was repeated for both the pre- and post-training sessions, and separately for the word list encoding and temporal order recognition tasks, as well as separately for the different groups in the athlete and training studies. The following statistics are identical to the above (For details on statistical analysis, see the section "Statistical analyses and thresholding").

Trainee-athlete RSA: degree of neural alignment between groups

To investigate whether method of loci training was associated with changes in the neural pattern similarity between the training study participants and memory athletes (or matched controls), we based on the beta-maps obtained from conventional across-trial modeling (thus, capturing average BOLD activity changes

during word list encoding and temporal order recognition). For each participant, neural activity was extracted from the pre-defined ROIs and the data were reshaped into a pattern vector for each ROI. The data were then *z*-scored across trials to remove mean activation differences. The pattern vectors of each ROI and each participant were then correlated to the pattern vectors of all other participants of the same group, using pairwise Pearson's correlation (*r*). The resulting correlations were Fisher's *z*-transformed. Trainee-athlete neural pattern similarity for each subject was then computed as the average pattern similarity of each participant in training group with those of each memory athlete, and, separately, with those of each individual of the matched control group (that is, the average correlation of voxel patterns from participant 1 with those of participants 1~*N* of athlete group or matched control group, separately). This procedure was repeated for both the pre- and post-training sessions, and separately for the word list encoding and temporal order recognition tasks, as well as separately for the different groups the training studies. This allowed us to compare the average neural pattern similarity between *time points* (pre- vs. post-training) and *groups* (neural pattern similarity between trainees × athletes vs. trainees × matched controls; Fig. 3a). The following equations were used to quantify the trainee-athlete RSA pattern similarity of individuals in the different experimental conditions. In brief, we calculated the pattern similarity of each participant in training study (here denoted as sub_1; to provide an example, we chose a participant of the memory training group, LOC) to the group average of that athlete group during the pre- (Equation 1) and post-training (Equation 2) sessions during, for example, word list encoding. Equations 3 and 4 follow the same rationale to assess the pattern similarity of the same sub_1 to the group average of that matched control (CON). In these equations, n denotes the total number of participants in the respective groups, and j represents the participant index for participants in athletes or matched control groups (For details on statistical analysis, see the section "Statistical analyses and thresholding").

$$S_{Pre}(Sub_1, LOC) = \frac{1}{n}\sum_{j=1}^{n} r_{Pre}\left(Sub_1, ATH_j\right) \quad (5)$$

$$S_{Post}(Sub_1, LOC) = \frac{1}{n}\sum_{j=1}^{n} r_{Post}\left(Sub_1, ATH_j\right) \quad (6)$$

$$S_{Pre}(Sub_1, LOC) = \frac{1}{n}\sum_{j=1}^{n} r_{Pre}\left(Sub_1, CON_j\right) \quad (7)$$

$$S_{Post}(Sub_1, LOC) = \frac{1}{n}\sum_{j=1}^{n} r_{Post}\left(Sub_1, CON_j\right) \quad (8)$$

Searchlight-based RSA

To go beyond the pre-defined ROIs and obtain whole-brain pattern similarity results, we repeated the above described RSA analyses but employed a searchlight with a radius of 6 mm (93 voxels) that was moved across the brain volume of each participant (Wagner, 2016; Wagner et al., 2020). Searchlights were considered only if they contained a minimum of 30 gray matter voxels. The average pattern similarity was calculated for each searchlight, and the resulting value was assigned to the center voxel of the respective searchlight. This process was then repeated until each voxel had served as the searchlight center once. The searchlight procedure was repeated for both pre- and post-training sessions, resulting in for 3-dimensional whole-brain pattern similarity maps (two for per task, two per session) per participant.

In the between-subject and trial-specific RSA analyses, we first examined the athlete study. Searchlight maps were subjected to independent-samples t-tests in a second-level group analysis to identify expertise-

related differences in neural representations, comparing memory athletes to matched controls. For the training study, paired-sample t-tests were conducted to assess training-related changes in neural representations between pre- and post-training sessions. Additionally, a one-way ANOVA was performed in the training study to directly compare neural representations across the three experimental groups (memory training group, active control group, and passive control group. Finally, in the trainee-athlete RSA analysis, searchlight maps were analyzed using a flexible 2 × 2 ANOVA design to investigate training-related changes in neural representations across the training study and the memory athlete study.

Prediction analysis

We then tested the behavioral relevance of changes in neural representations between all individuals of the training study and memory athletes with respect to long-term memory performance. A balanced 4-fold cross-validation procedure (Cohen, 2010; Supekar et al., 2013) was employed to predict the improvement in free recall performance resulting from method of loci training. Specifically, we modeled the change in free recall performance from baseline to two long-term re-test intervals: 24 hours, and 4 months. These predictions were based on changes in neural pattern similarity between all individuals of the training study and athletes from pre- to post-training. For each model included one dependent variable, such as free recall performance improvement at the 24-hour re-test (performance at the 24-hour post-training session minus the 24-hour pre-training session), or the 4-month re-test (performance at 4 months minus the 20-minute pre-training session). The independent variables differed by task: i) word list encoding: the increased distinct neural representations between all individuals of the training study and memory athletes (i.e., neural pattern similarity pre-training minus post-training) were used to test the behavioral relevance of distinct neural representations with respect to long-term memory performance. ii) temporal order recognition: the increased shared neural representations between all individuals of the training study and memory athletes (i.e., neural pattern similarity post-training minus pre-training) were used to evaluate the behavioral relevance of shared neural representations among experienced individuals in relation to long-term memory performance.

The relationship between the independent and dependent variables was quantified as $r_{(\text{predicted,observed})}$, measuring the degree to which the independent variables predicted the dependent variable. This measure was derived using the balanced 4-fold cross-validation procedure, in which the data for all individuals in the training study were divided into four equal subsets while ensuring that the distributions of dependent and independent variables were balanced across folds. This means that to validate the robustness of the relationship between training-related changes in participants' neural representations and individual memory performance, we tested the model's generalizability to out-of-sample individuals. In each iteration, the linear regression model was trained on three folds, and the held-out fold was used for prediction. This process was repeated until each fold served as the prediction sample once. The final $r_{(\text{predicted, observed})}$ value represented the average correlation between predicted and observed data across all folds.

To assess statistical significance, we employed a nonparametric approach. For instance, under the null hypothesis, no significant relationship would exist between free recall performance after 4 months (dependent variable) and trainee-athlete neural representation alterations (independent variable) in the specified ROIs. We generated 1000 surrogate datasets with randomly shuffled participant labels that were denoted as D_i. These surrogate datasets were created by permuting the labels assigned to the observed data points. The $r_{(\text{predicted, observed})i}$ was computed by using the true labels of D*i* and the predicted labels generated through the 4-fold cross-validation procedure. This yielded a null distribution of $r_{(\text{predicted, observed})}$ (Cohen, 2010). The *p*-value was determined by counting the number of times when $r_{(\text{predicted, observed})i}$ was greater than $r_{(\text{predicted, observed})}$, divided that count by the number of D*i* datasets ($[r_{(\text{predicted, observed})i} > r_{(\text{predicted, observed})}]/Di$, D*i* is 1000 in our case)(Qin et al., 2014; Zhu et al., 2022).

Statistical analyses and thresholding

Statistical analysis of behavioral data and ROI-based RSA results

Analyses of ROI-based RSA results (including between-subject, trial-specific and trainee-athlete RSA) were conducted using jamovi (version 2.3.13.0; https://www.jamovi.org/). For the between-subject and trial-specific RSA, data from the athlete study (comparing memory athletes to matched controls) were subjected to independent-samples *t*-tests to evaluate expertise-related differences in neural representations. In the training study, paired-sample *t*-tests were used to examine training-related changes in neural representation between pre- and post-training sessions. For the trainee-athlete RSA, the average neural pattern similarity was compared across time points (pre- vs. post-training) and groups (neural pattern similarity between trainees × athletes vs. trainees × matched controls). These comparisons were further analyzed using a 2 × 2 repeated-measures ANOVA to assess training-related changes in neural representations between the training study and the athlete study. Post-hoc pairwise comparisons following significant ANOVA effects were adjusted for multiple comparisons using the Bonferroni correction. Effect sizes were reported as Cohen's d for *t*-tests and eta squared (η^2) for ANOVA models (Lakens, 2013). To account for multiple comparisons across the four selected regions of interest (ROIs), false discovery rate (FDR) correction using the Benjamini–Hochberg procedure was applied at a threshold of $q < 0.05$. This correction was conducted separately for the word list encoding task and the temporal order recognition task. For all analyses, the significance level (α) was set at 0.05 (two-tailed).

Statistical thresholding of whole-brain RSA results and anatomical labeling

Unless stated otherwise, significance for all whole-brain fMRI analyses was assessed using cluster-based inference with a cluster-defining threshold of $p < 0.001$ and a cluster probability of $p < 0.05$ family-wise error (FWE) corrected for multiple comparisons. The corrected cluster size (i.e., the spatial extent of a cluster that is required to be labeled as significant) was calculated using the SPM extension "CorrClusTh.m" and the Newton-Raphson search method (script provided by Thomas Nichols, University of Warwick, United Kingdom, and Marko Wilke, University of Tübingen, Germany; http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/nichols/scripts/spm/). Anatomical nomenclature for all tables was obtained from the Laboratory for Neuro Imaging (LONI) Brain Atlas (LBPA40, www. loni.usc.edu/atlases/).

Data availability

Raw, anonymized fMRI data are available upon reasonable request to the corresponding authors (martin.dresler@donders.ru.nl, isabella.wagner@univie.ac.at). At present, participant informed consent does not allow for depositing the full data set. All analyses are based on openly available software (see section "Code availability" below). Source data to reproduce all graphs (behavioral performance and ROI-based results of all analyses), as well as (un-)thresholded statistical whole-brain fMRI maps and their corresponding results tables, are openly available at the Open Science Framework (https://osf.io/7hb53/).

Code availability

All analyses are based on openly available software or custom code that is provided on the Open Science Framework (https://osf.io/7hb53/).

References

Amer, T., & Davachi, L. (2023). Extra-hippocampal contributions to pattern separation. *eLife*, *12*, e82250. https://doi.org/10.7554/eLife.82250

Baddeley, A., Conway, M., Aggleton, J., Burgess, N., Becker, S., King, J. A., & O'Keefe, J. (2001). Memory for events and their spatial context: Models and experiments. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356*(1413), 1493–1503. https://doi.org/10.1098/rstb.2001.0948

Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, *45*(13), 2883–2901. https://doi.org/10.1016/j.neuropsychologia.2007.06.015

Bellmund, J. L. S., Deuker, L., Montijn, N. D., & Doeller, C. F. (2022). Mnemonic construction and representation of temporal structure in the hippocampal formation. *Nature Communications*, *13*(1), 3395. https://doi.org/10.1038/s41467-022-30984-3

Blumenfeld, R. S., & Ranganath, C. (2007). Prefrontal Cortex and Long-Term Memory Encoding: An Integrative Review of Findings from Neuropsychology and Neuroimaging. *The Neuroscientist*, *13*(3), 280–291. https://doi.org/10.1177/1073858407299290

Boisgueheneuc, F. du, Levy, R., Volle, E., Seassau, M., Duffau, H., Kinkingnehun, S., Samson, Y., Zhang, S., & Dubois, B. (2006). Functions of the left superior frontal gyrus in humans: A lesion study. *Brain*, *129*(12), 3315–3328. https://doi.org/10.1093/brain/awl244

Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The Human Hippocampus and Spatial and Episodic Memory. *Neuron*, *35*(4), 625–641. https://doi.org/10.1016/S0896-6273(02)00830-9

Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D., & Kriegeskorte, N. (2014). Unique semantic space in the brain of each beholder predicts perceived similarity. *Proceedings of the National Academy of Sciences*, *111*(40), 14565–14570. https://doi.org/10.1073/pnas.1402594111

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017a). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, *20*(1), 115–125. https://doi.org/10.1038/nn.4450

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017b). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, *20*(1), Article 1. https://doi.org/10.1038/nn.4450

Cohen, J. R. (2010). Decoding developmental differences and individual variability in response inhibition through predictive analyses across individuals. *Frontiers in Human Neuroscience*, *4*. https://doi.org/10.3389/fnhum.2010.00047

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, *15*(6), 254–262. https://doi.org/10.1016/j.tics.2011.04.003

Dijkstra, N., Bosch, S. E., & Gerven, M. A. J. van. (2019). Shared Neural Mechanisms of Visual Perception and Imagery. *Trends in Cognitive Sciences*, *23*(5), 423–434. https://doi.org/10.1016/j.tics.2019.02.004

Dresler, M., Shirer, W. R., Konrad, B. N., Müller, N. C. J., Wagner, I. C., Fernández, G., Czisch, M., & Greicius, M. D. (2017). Mnemonic Training Reshapes Brain Networks to Support Superior Memory. *Neuron*, *93*(5), 1227-1235.e6. https://doi.org/10.1016/j.neuron.2017.02.003

Duszkiewicz, A. J., McNamara, C. G., Takeuchi, T., & Genzel, L. (2019). Novelty and Dopaminergic Modulation of Memory Persistence: A Tale of Two Systems. *Trends in Neurosciences*, *42*(2), 102–114. https://doi.org/10.1016/j.tins.2018.10.002

Fernández, G., & Morris, R. G. M. (2018). Memory, Novelty and Prior Knowledge. *Trends in Neurosciences*, *41*(10), 654–659. https://doi.org/10.1016/j.tins.2018.08.006

Ghazizadeh, A., Hong, S., & Hikosaka, O. (2018). Prefrontal Cortex Represents Long-Term Memory of Object Values for Months. *Current Biology*, *28*(14), 2206-2217.e5. https://doi.org/10.1016/j.cub.2018.05.017

Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nature Neuroscience*, *19*(5), 665–667. https://doi.org/10.1038/nn.4284

Horner, A. J., Bisby, J. A., Bush, D., Lin, W.-J., & Burgess, N. (2015). Evidence for holistic episodic recollection via hippocampal pattern completion. *Nature Communications*, *6*(1), 7462. https://doi.org/10.1038/ncomms8462

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(19), 6829–6833. https://doi.org/10.1073/pnas.0801268105

Kishiyama, M. M., Yonelinas, A. P., & Knight, R. T. (2009). Novelty Enhancements in Memory Are Dependent on Lateral Prefrontal Cortex. *Journal of Neuroscience*, *29*(25), 8114–8118. https://doi.org/10.1523/JNEUROSCI.5507-08.2009

Koch, G. E., Paulus, J. P., & Coutanche, M. N. (2020). Neural Patterns are More Similar across Individuals during Successful Memory Encoding than during Failed Memory Encoding. *Cerebral Cortex*, *30*(7), 3872–3883. https://doi.org/10.1093/cercor/bhaa003

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The Architecture of Cognitive Control in the Human Prefrontal Cortex. *Science*, *302*(5648), 1181–1185. https://doi.org/10.1126/science.1088545

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4. https://doi.org/10.3389/neuro.06.004.2008

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00863

LaRocque, K. F., Smith, M. E., Carr, V. A., Witthoft, N., Grill-Spector, K., & Wagner, A. D. (2013). Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(13), 5466–5474. https://doi.org/10.1523/JNEUROSCI.4293-12.2013

Liu, C., Ye, Z., Chen, C., Axmacher, N., & Xue, G. (2022). Hippocampal Representations of Event Structure and Temporal Context during Episodic Temporal Order Memory. *Cerebral Cortex*, *32*(7), 1520–1534. https://doi.org/10.1093/cercor/bhab304

Lyu, D., Stieger, J. R., Xin, C., Ma, E., Lusk, Z., Aparicio, M. K., Werbaneth, K., Perry, C. M., Deisseroth, K., Buch, V., & Parvizi, J. (2023). Causal evidence for the processing of bodily self in the anterior precuneus. *Neuron*, *111*(16), 2502-2512.e4. https://doi.org/10.1016/j.neuron.2023.05.013

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, *503*(7474), 78–84. https://doi.org/10.1038/nature12742

Mason, R. A., & Just, M. A. (2016). Neural Representations of Physics Concepts. *Psychological Science*, *27*(6), 904–913. https://doi.org/10.1177/0956797616641941

Matsumoto, M., Matsumoto, K., & Tanaka, K. (2007). Effects of novelty on activity of lateral and medial prefrontal neurons. *Neuroscience Research*, *57*(2), 268–276. https://doi.org/10.1016/j.neures.2006.10.017

Meshulam, M., Hasenfratz, L., Hillman, H., Liu, Y.-F., Nguyen, M., Norman, K. A., & Hasson, U. (2021). Neural alignment predicts learning outcomes in students taking an introduction to computer science course. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-021-22202-3

Müller, N. C. J., Konrad, B. N., Kohn, N., Muñoz-López, M., Czisch, M., Fernández, G., & Dresler, M. (2018). Hippocampal-caudate nucleus interactions support exceptional memory performance. *Brain Structure & Function*, *223*(3), 1379–1389. https://doi.org/10.1007/s00429-017-1556-2

Mumford, J. A., Davis, T., & Poldrack, R. A. (2014). The impact of study design on pattern estimation for single-trial multivariate pattern analysis. *NeuroImage*, *103*, 130–138. https://doi.org/10.1016/j.neuroimage.2014.09.026

Nash, M. I., Hodges, C. B., Muncy, N. M., & Kirwan, C. B. (2021). Pattern separation beyond the hippocampus: A high-resolution whole-brain investigation of mnemonic discrimination in healthy adults. *Hippocampus*, *31*(4), 408–421. https://doi.org/10.1002/hipo.23299

Nguyen, M., Chang, A., Micciche, E., Meshulam, M., Nastase, S. A., & Hasson, U. (2022). Teacher–student neural coupling during teaching and learning. *Social Cognitive and Affective Neuroscience*, *17*(4), 367–376. https://doi.org/10.1093/scan/nsab103

Oedekoven, C. S. H., Keidel, J. L., Berens, S. C., & Bird, C. M. (2017). Reinstatement of memory representations for lifelike events over the course of a week. *Scientific Reports*, *7*(1), 14305. https://doi.org/10.1038/s41598-017-13938-4

Op de Beeck, H. P., Haushofer, J., & Kanwisher, N. G. (2008). Interpreting fMRI data: Maps, modules and dimensions. *Nature Reviews Neuroscience*, *9*(2), 123–135. https://doi.org/10.1038/nrn2314

Owen, A. M. (2000). The role of the lateral frontal cortex in mnemonic processing: The contribution of functional neuroimaging. *Experimental Brain Research*, *133*(1), 33–43. https://doi.org/10.1007/s002210000398

Öztekin, I., Curtis, C. E., & McElree, B. (2009). The Medial Temporal Lobe and the Left Inferior Prefrontal Cortex Jointly Support Interference Resolution in Verbal Working Memory. *Journal of Cognitive Neuroscience*, *21*(10), 1967–1979. https://doi.org/10.1162/jocn.2008.21146

Pacheco Estefan, D., Sánchez-Fibla, M., Duff, A., Principe, A., Rocamora, R., Zhang, H., Axmacher, N., & Verschure, P. F. M. J. (2019). Coordinated representational reinstatement in the human hippocampus and lateral temporal cortex during episodic memory retrieval. *Nature Communications*, *10*(1), 2255. https://doi.org/10.1038/s41467-019-09569-0

Parker, A., Wilding, E., & Akerman, C. (1998). The Von Restorff effect in visual object recognition memory in humans and monkeys. The role of frontal/perirhinal interaction. *Journal of Cognitive Neuroscience*, *10*(6), 691–703. https://doi.org/10.1162/089892998563103

Petrides, M. (2000). The role of the mid-dorsolateral prefrontal cortex in working memory. *Experimental Brain Research*, *133*(1), 44–54. https://doi.org/10.1007/s002210000399

Qin, S., Cho, S., Chen, T., Rosenberg-Lee, M., Geary, D. C., & Menon, V. (2014). Hippocampal-neocortical functional reorganization underlies children's cognitive development. *Nature Neuroscience*, *17*(9), Article 9. https://doi.org/10.1038/nn.3788

Ranganath, C., & Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nature Reviews Neuroscience*, *4*(3), 193–202. https://doi.org/10.1038/nrn1052

Ranganath, C., & Ritchey, M. (2012). Two cortical systems for memory-guided behaviour. *Nature Reviews Neuroscience*, *13*(10), 713–726. https://doi.org/10.1038/nrn3338

Sakai, K., Rowe, J. B., & Passingham, R. E. (2002). Active maintenance in prefrontal area 46 creates distractor-resistant memory. *Nature Neuroscience*, *5*(5), 479–484. https://doi.org/10.1038/nn846

Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature Communications*, *6*(1), Article 1. https://doi.org/10.1038/ncomms9151

Sereno, M. I., Pitzalis, S., & Martinez, A. (2001). Mapping of Contralateral Space in Retinotopic Coordinates by a Parietal Cortical Area in Humans. *Science*, *294*(5545), 1350–1354. https://doi.org/10.1126/science.1063695

Sestieri, C., Shulman, G. L., & Corbetta, M. (2017). The contribution of the human posterior parietal cortex to episodic memory. *Nature Reviews Neuroscience*, *18*(3), 183–192. https://doi.org/10.1038/nrn.2017.6

Sheng, J., Wang, S., Zhang, L., Liu, C., Shi, L., Zhou, Y., Hu, H., Chen, C., & Xue, G. (2023). Intersubject similarity in neural representations underlies shared episodic memory content. *Proceedings of the National Academy of Sciences*, *120*(35), e2308951120. https://doi.org/10.1073/pnas.2308951120

Stevenson, R. F., Reagh, Z. M., Chun, A. P., Murray, E. A., & Yassa, M. A. (2020). Pattern Separation and Source Memory Engage Distinct Hippocampal and Neocortical Regions during Retrieval. *Journal of Neuroscience*, *40*(4), 843–851. https://doi.org/10.1523/JNEUROSCI.0564-19.2019

Takashima, A., Nieuwenhuis, I. L. C., Rijpkema, M., Petersson, K. M., Jensen, O., & Fernández, G. (2007). Memory trace stabilization leads to large-scale changes in the retrieval network: A functional MRI study on associative memory. *Learning & Memory*, *14*(7), 472–479. https://doi.org/10.1101/lm.605607

Tanguay, A. F., Palombo, D. J., Love, B., Glikstein, R., Davidson, P. S., & Renoult, L. (2023). The shared and unique neural correlates of personal semantic, general semantic, and episodic memory. *eLife*, *12*, e83645. https://doi.org/10.7554/eLife.83645

Wagner, I. C. (2016). The Integration of Distributed Memory Traces. *The Journal of Neuroscience*, *36*(42), 10723–10725. https://doi.org/10.1523/JNEUROSCI.2344-16.2016

Wagner, I. C., Konrad, B. N., Schuster, P., Weisig, S., Repantis, D., Ohla, K., Kühn, S., Fernández, G., Steiger, A., Lamm, C., Czisch, M., & Dresler, M. (2021). Durable memories and efficient neural coding through mnemonic training using the method of loci. *Science Advances*, *7*(10), eabc7606. https://doi.org/10.1126/sciadv.abc7606

Wagner, I. C., Rütgen, M., & Lamm, C. (2020). Pattern similarity and connectivity of hippocampal-neocortical regions support empathy for pain. *Social Cognitive and Affective Neuroscience*, *15*(3), 273–284. https://doi.org/10.1093/scan/nsaa045

Xiao, X., Zhou, Y., Liu, J., Ye, Z., Yao, L., Zhang, J., Chen, C., & Xue, G. (2020). Individual-specific and shared representations during episodic memory encoding and retrieval. *NeuroImage*, *217*, 116909. https://doi.org/10.1016/j.neuroimage.2020.116909

Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*(10), 515–525. https://doi.org/10.1016/j.tins.2011.06.006

Yates, F. A. (1966). *The Art of Memory*. Routledge and Kegan Paul.

Yeshurun, Y., Swanson, S., Simony, E., Chen, J., Lazaridi, C., Honey, C. J., & Hasson, U. (2017). Same Story, Different Story: The Neural Representation of Interpretive Frameworks. *Psychological Science*, *28*(3), 307–319. https://doi.org/10.1177/0956797616682029

Zhu, Y., Zeng, Y., Ren, J., Zhang, L., Chen, C., Fernandez, G., & Qin, S. (2022). Emotional learning retroactively promotes memory integration through rapid neural reactivation and reorganization. *eLife*, *11*, e60190. https://doi.org/10.7554/eLife.60190

Zsido, R. G., Williams, A. N., Barth, C., Serio, B., Kurth, L., Mildner, T., Trampel, R., Beyer, F., Witte, A. V., Villringer, A., & Sacher, J. (2023). Ultra-high-field 7T MRI reveals changes in human medial temporal lobe volume in female adults during menstrual cycle. *Nature Mental Health*, *1*(10), Article 10. https://doi.org/10.1038/s44220-023-00125-w