nature human behaviour

Registered Report

https://doi.org/10.1038/s41562-025-02390-4

Neural activations and representations during episodic versus semantic memory retrieval

Received: 10 July 2020

Accepted: 3 December 2025

Published online: 27 January 2026

Check for updates

Roni Tibon[1,2] ✉, Andrea Greve[2], Gina Humphreys[2], Jörn Alexander Quent[2,3,4] & Richard Henson[2,5]

Although the distinction between episodic and semantic memory is supported by numerous neuropsychological studies, neuroimaging data have shown considerable overlap between regions that are activated during semantic and episodic remembering. This might indicate similar or shared mechanisms but might also result from inadequate task designs or poor functional magnetic resonance imaging signal coverage. Here we compared neural activations and representations associated with successful retrieval of episodic and semantic memories, using tasks that are more closely matched. A total of $n = 40$ participants recalled pairings between logos and brand names, where the pairings corresponded to real-world knowledge (semantic task) or were learned in an initial study phase (episodic task). Neither a priori-defined networks nor clusters generally activated by our task provided evidence for any difference between successful semantic and episodic retrieval, with the Bayes factor for the a priori networks supporting the null hypothesis of no difference. Protocol registration The Stage 1 protocol for this Registered Report was accepted in principle on 15 September 2021. The protocol, as accepted by the journal, can be found at https://osf.io/dm47y/.

Recently, there has been renewed interest in the long proposed distinction between two different types of declarative memory: episodic and semantic[1]. Episodic memory refers to the ability to recollect a past event that occurred in a particular spatial and temporal context. This type of memory supports the human capacity to re-experience events from our past, as a form of 'mental time travel'[2]. Semantic memory, on the other hand, refers to the ability to remember facts and general knowledge about the world that are retrieved independently from their original spatial or temporal context.

In many ways, this distinction has dominated research on the cognitive neuroscience of memory and remains of central importance to the field. It is supported by a large corpus of neuropsychological studies, including studies of neurodegenerative disorders and selective brain lesions. For example, early stages of Alzheimer's disease, which are characterized mainly by medial temporal lobe (MTL) degeneration, typically produce deficits in episodic memory, whereas semantic memory remains intact until later stages of the disease[3,4] (see also findings of intact ability to acquire new semantic knowledge in developmental amnesia[5]). Conversely, semantic dementia, a younger-onset neurodegenerative disorder that is associated with degeneration that starts in the anterior temporal lobe (ATL), produces a multimodal loss of semantic knowledge, while episodic memory appears relatively spared until later in the disease[6–9].

Despite this historical distinction between episodic and semantic processing, established via neuropsychological and clinical studies,

[1]School of Psychology, University of Nottingham, Nottingham, UK. [2]MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK. [3]Institute of Science and Technology for Brain-inspired Intelligence, Fudan University, Shanghai, China. [4]MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China. [5]Department of Psychiatry, University of Cambridge, Cambridge, UK. ✉e-mail: roni.tibon@nottingham.ac.uk

the picture arising from neuroimaging studies is much less conclusive. Indeed, neuroimaging studies with healthy human participants seem to challenge the episodic–semantic dissociation, by revealing considerable overlap in the brain regions involved in semantic and episodic processing[10]. Episodic memory has been associated with enhanced activity in a consistent set of brain regions, termed the 'core recollection network', regardless the nature of the recollected content[11,12]. Similarly, semantic memory has been associated with a core network termed 'the general semantic network' that is responsive to tasks that require semantic memory retrieval, for example, word versus non-word processing[13]. Importantly, these two networks have extensive overlap, specifically in parahippocampal, middle temporal, inferior parietal, posterior cingulate/precuneus and midline frontal regions[10]. These findings inspired alternative perspectives, for example, that episodic and semantic memories lie along a continuum of contextualization within time, space and valence that is processed by a unitary memory system[10] or that the dichotomy between the systems should be replaced with a continuum spanning from concrete to abstract representations[14]. Such proposals, which have received increasing attention in recent years, further prompt the reconsideration of evidence from clinical populations, suggesting that the boundaries between the two memory systems might be unclear[15].

Because, for many years, episodic and semantic memories were considered to be distinct entities, a research tradition had developed in which they are explored separately. The consequence of this is a lack of within-study designs that tap into both systems, which, ironically, prevents adequate rebuttal to the unifying theories described above. Although there have been reviews and meta-analyses, these cannot provide conclusive evidence either: anatomical overlap might lead to the conclusion that there is a functional overlap, but might also be the result of blurring small-scale variability across studies. Similarly, a spatial segregation might lead to the conclusion that the systems are distinct but may also result from experimental or procedural variations such as different scanning parameters designed to maximize ATL or MTL signal for semantic versus episodic tasks. Therefore, to provide compelling evidence supporting one view or the other, within-study (and preferably, within-participant) comparisons are crucial. The few studies that have compared episodic and semantic memory within the same experimental design have tended to observe extensive overlap of functional networks/brain regions[16–22] (though others have found significant differences between the networks engaged[23]). Although this overlap might reflect shared mechanisms or even a single common memory system, current data do not permit a decisive conclusion for several reasons. First, this overlap could also result from experimental designs in which semantic and episodic processes are not fully dissociated. For example, in studies comparing averaged activity during episodic and semantic tasks common activation might reflect shared general processes (working memory, perceptual processes, attention and so on) rather than being specific to episodic or semantic retrieval, even when only including successful trials[16,17]. Furthermore, the use of recognition memory tasks to measure episodic retrieval[18,24] allows contributions from familiarity-based judgements that are arguably more 'semantic' than recollection-based judgements, as they too lack spatiotemporal context. Second, the degree of overlap between episodic and semantic tasks may have been exaggerated because of limitations with typical functional magnetic resonance imaging (fMRI) protocols. Namely, the ATL is considered to be a core multimodal semantic hub that is essential for semantic, but not episodic, retrieval[25], yet typical fMRI protocols have poor signal coverage in this region. Third, while previous studies have compared differences in mean brain activity during episodic and semantic processing, to our knowledge, no previous study has directly compared multivoxel patterns in these core systems, which can potentially capture differences in the representations of retrieved episodic versus semantic memories (one study compared multivoxel patterns of episodic and sematic memories but only within the angular gyrus and sensory areas[26]).

In the current fMRI study, we aimed to distinguish retrieval of semantic and episodic memories using closely matched tasks by using a pair-associate task that compares episodic and semantic retrieval triggered by the same cue. In this task, schematically depicted in Fig. 1, episodic and semantic processes are minimally confounded as cued-recall is used to preclude familiarity-based judgements. In addition, an optimized multi-echo protocol was used to provide whole-brain fMRI coverage. Participants completed two critical tasks—episodic and semantic—on two separate days, as well as an additional control (baseline) task. The stimulus pairs used in both the semantic and episodic tasks consisted of familiar and unfamiliar logos (pictures) and names (words) of brands (for example, 'google'). In the semantic task, these pairings matched those in the real world, though the logos were chosen so that a typical participant only knew approximately one half of them. In the episodic task, the pairings were established instead in an initial unscanned study phase, where a logo was presented together with the name of a different brand. During the subsequent test phases, fMRI data were locked to the presentation of the cue, while participants were trying to recall details about the associated brand from their prior knowledge in the semantic task or details about the pairing in the study phase in the episodic task and subsequently indicate the level of detail of their memory. This cued-recall procedure allowed us to prevent guessing. Moreover, with our design, episodic information might help retrieve the semantic associate but not without actual semantic knowledge of the pairing. Similarly, semantic knowledge might facilitate episodic retrieval, but not without actual episodic knowledge. Thus, even if the alternative process (that is, semantic memory in the episodic task and episodic memory in the semantic task) might provide some memory scaffolding in the same way that attention or visual processing can support memory, it could not solely determine success. In other words, a correct response cannot be provided without information from the probed process. Two test phases were employed. In the first test phase, the cue was the logo. The data from this test phase were used for the univariate analysis. In the second test phase, the cue was the name of the brand. The correspondence between data from both test phases was used in the multivoxel similarity analysis (see below).

Rather than comparing averaged activity for each task, we focused on the difference between successful versus failed retrieval trials within each task, which is more closely controlled. More specifically, we compared successful versus failed retrieval when cued by the logo in the first test phase and examined whether this difference implicates different brain regions for episodic and semantic recall. Moreover, in addition to mean activity over voxels within brain regions, we also looked at multivoxel patterns that are likely to reflect the content of what had been retrieved. For the latter, we correlated activity patterns elicited when the logo served as the cue in the first test phase, on trials when its brand was successfully recalled, with activity patterns elicited when the brand's name served as the cue in the second test phase. We compared this correlation with the correlation between the pattern elicited by the logo and the pattern that was elicited when an unassociated brand name was presented. If activity patterns elicited by logo cues are more similar to those elicited by their corresponding name than to those elicited by other 'unassociated' names, then this constitutes evidence for content-specific representations. We then examined whether this difference in correlated activity patterns was associated with different brain regions for semantic and episodic memories.

Our hypotheses are detailed in Table 1. We hypothesized that the brain regions (Fig. 2) that support and represent successful episodic retrieval are distinct from those that support and represent successful semantic retrieval. Therefore, when contrasting successful versus failed retrieval, we predicted greater overall activation in episodic brain regions for the episodic task compared with the semantic and control tasks but greater overall activation in semantic brain regions for the semantic task compared with the episodic and control tasks. Similarly, when correlating multivoxel patterns, we predicted a greater similarity

https://doi.org/10.1038/s41562-025-02390-4

Table 1 | Design table

Question	Hypothesis	Sampling plan (for example, power analysis)	Analysis plan	Interpretation given to different outcomes
Can we distinguish successful episodic retrieval from semantic retrieval?	The brain regions supporting successful episodic retrieval are distinct from those supporting successful semantic retrieval	Final sample size will be determined using a Bayesian sequential design with maximal n=100, as detailed in the 'Participants' section and Supplementary Analysis 1	fMRI univariate analysis will be conducted for single-trial beta values for logo-cue trials classified as 'success' versus 'failure' trials, in the semantic versus episodic task. For each ROI, we will fit a Bayesian mixed-effect model with the formula: betas ≈ ROIs × task × response type + (1 ROIs + task + response type|participant). We will then test our prediction of greater recall success effect in the episodic task in episodic ROIs, but in the semantic task in semantic ROIs. Such pattern will support our hypothesis that the mechanisms supporting episodic and semantic retrieval are dissociable, as they are supported by different brain regions	Using a Bayesian inference, a BF > 10 would support our hypothesis that episodic and semantic memories are dissociable and are supported by different neural mechanisms. By contrast, if the null hypothesis is supported, this would indicate that episodic and semantic memories are processed in a similar fashion. This would provide support to the view of a shared mechanisms or a single common system for episodic and semantic memories
Can we distinguish content representation of episodic and semantic memories?	The brain regions that represent episodic content are distinguished from those representing semantic content		fMRI similarity analysis will be conducted for the correlation coefficient of success logo-cue trials when correlated with a name-cue trial representing the same episodic/semantic instant, versus a name-cue trial representing a different instant. For each ROI, we will fit a Bayesian mixed-effect model with the formula: correlation ≈ ROIs × task × trial type + (1 + ROIs + task + trial type|participant). We will then test our prediction of greater similarity effect in the episodic task in episodic ROIs but in the semantic task in semantic ROIs. Such pattern will support our hypothesis episodic and semantic memories are dissociable, as they are represented in different brain regions	BF > 10 would support our hypothesis that episodic and semantic memories are represented within different brain regions. By contrast, if the null hypothesis is supported, this would indicate that episodic and semantic memories are represented within the same brain regions

effect in the episodic task in the episodic brain regions but a greater similarity effect in the semantic task in the semantic brain regions.

Results

Behavioural analyses

Figure 3 shows the distribution of accuracy rates across the three tasks. As can be seen, although success rates were not at 50%, they were not too far from this, leaving us with a sufficient number of successful and unsuccessful trials for all participants. To examine the distribution of success and failure trials across the different tasks, we ran two Bayesian logistic regression models. The first model (m1) included task (episodic, semantic and control) as a fixed effect and a participant-specific intercept and slope as random effects. The second model (m2) was similar but did not include task as a fixed factor. This model comparison targets the effect of task, such that if accuracy differs across the different tasks, then the Bayes factor (BF) should favour m1 (which includes the task factor) over m2, thus indicating support for H_1. The analysis favoured H_1 over the null hypothesis (BF_{10} = 1,780.76), suggesting that accuracy rates differed across tasks.

Exploratory behavioural analyses

Comparing performance in critical tasks. As a follow-up analysis, we reran the analysis described above but only included data from the two critical tasks (episodic, semantic). This analysis showed that participants' performance differed between the tasks (BF = 1,894.41), such that accuracy rates were greater in the episodic versus semantic task. Although differences in accuracy could confound fMRI comparisons in analyses that collapse over success and fail trials, note that in the current study responses to individual trials were compared as a function of their success/fail status, which should be independent of overall numbers of each trial type, particularly since guesses are unlikely to contaminate our definition of success/fail.

Exploring differences in 'detailedness'. This exploratory analysis was performed to account for the possibility that some of the neural differences that we have observed are due to variations in vividness between the tasks (Discussion). To this end, we compared participants' reports of recalled 'detailedness' (as a proxy of vividness) by contrasting the proportion of trials for which they have reported that many details were recalled in the episodic task (mean of 0.82, s.d. of 0.18) and semantic task (mean of 0.77, s.d. of 0.17). A paired-sample Bayesian t-test did not reveal any conclusive evidence for a difference between the tasks, BF_{01} = 2.34.

Univariate analyses

A priori-defined networks. Beta values for recall success and failure trials across tasks and networks are shown in Fig. 4a, left. To test our prediction of greater recall success effect in the episodic task in the episodic network, but greater recall success effect in the semantic task in the semantic network, we used Bayesian linear mixed-models, as described above.

We first ran this analysis after excluding trials corresponding to logos that were classified as 'can name' in the debriefing session, to minimize potential influences of semantic processing in the episodic task. Contrary to our prediction, this analysis favoured H_0 over the alternative hypothesis (BF_{01} = 12.67), suggesting that the null hypothesis of no difference in success effects across networks is preferred. We then reinstated the trials that were excluded and ran the analysis again. This analysis again favoured H_0 over the alternative hypothesis (BF_{01} = 13.33), suggesting that the null hypothesis is again preferred. Importantly, as this analysis determined our stopping criteria (Table 1), no additional data were collected.

Data-defined clusters. Our preregistered analysis of data-defined clusters yielded two large clusters: one that included much of the cerebral cortex and another that included much of the cerebellum (Fig. 2). Beta values, averaged across all voxels in these clusters, for recall success and failure trials in each task and each cluster are shown in Fig. 4b, left. We ran the same analysis used for the a priori-defined networks. With this analysis, H_0 was favoured over the alternative (BF_{01} = 3.33), but this did not reach our prespecified criteria of BF > 10, and therefore, this result is considered inconclusive.

Examination of an overall task effect. In our Stage 1 report, we noted that to facilitate the interpretation of the univariate results, we will test for overall task effects (ignoring whether trials were successful or not)

https://doi.org/10.1038/s41562-025-02390-4

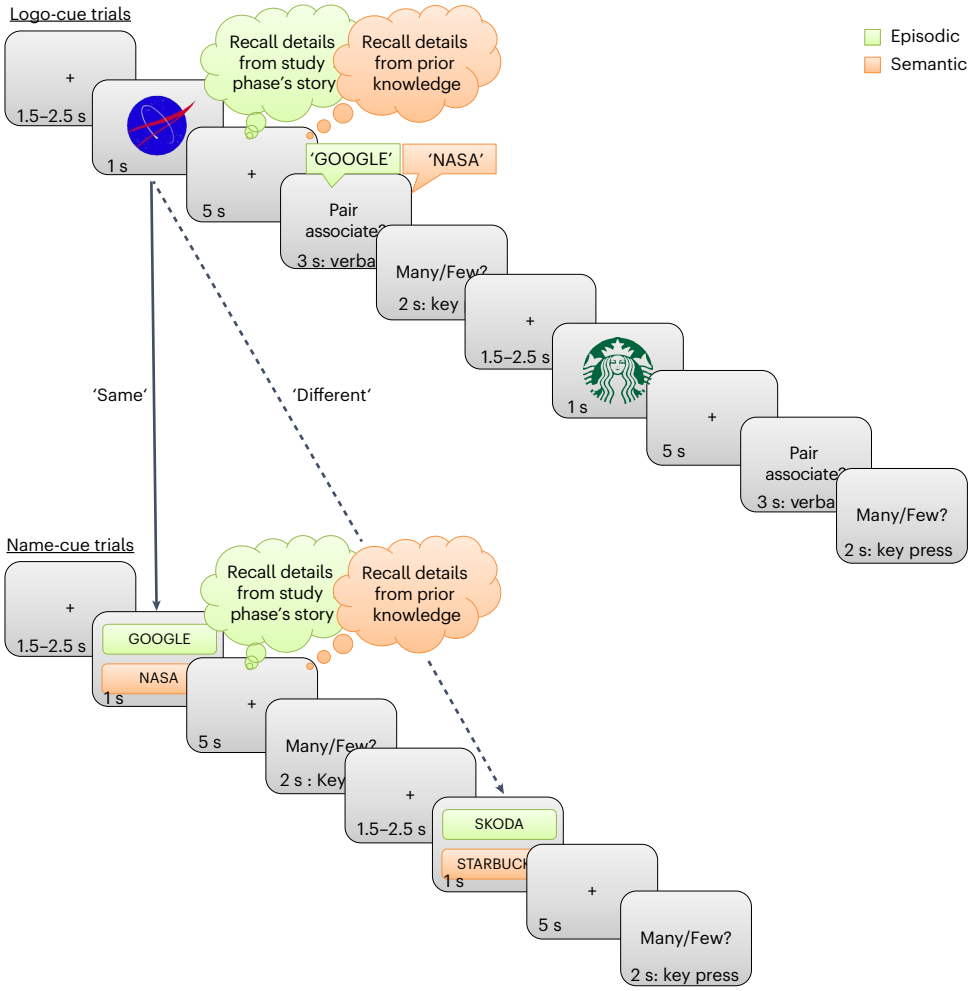


Fig. 1 | Schematic illustration of the test phase of the two critical tasks. Top row: during logo-cue trials, a picture of a logo was presented as the cue. Participants had to recall the associated brand from the study phase (in the episodic task; green shades) or from their prior knowledge (in the semantic task; orange shades) and to indicate whether they remember many or few associated details. Bottom row: during name-cue trials, a brand's name was presented, and participants indicated whether they could think of many or few associated details. For the univariate analysis, difference in activity for successful logo-cue (locked to the presentation of the cue) versus failure logo-cue trials in the episodic task was contrasted with that difference in the semantic task, to test for overall episodic/semantic recall success effects. For the similarity analysis, patterns of activity elicited by successful logo-cue trials (during the 5 s delay period) were correlated with patterns of activity elicited by the corresponding name-cue trials ('same'; solid line) and with patterns elicited by non-corresponding name-cue trials ('different'; dashed line). The difference between these correlations in the episodic task was then contrasted with that difference in the semantic task, to test for episodic/semantic pattern-specific effects.

in a priori-defined networks. General activation related to retrieval mode or retrieval orientation[27] would predict greater overall activation for both success and failure trials in the episodic network for the episodic task, but greater overall activation in the semantic network for the semantic task. We used the Savage–Dickey density ratio method[28] to calculate the BF for the hypothesized task × region of interest (ROI) interaction, but this did not show conclusive support for our hypothesis ($BF_{10} = 2.50$). Nonetheless, given the behavioural differences between the tasks with better and more variable performance in the episodic compared with the semantic task, analyses such as this that do not take the recall success contrast into account are hard to interpret.

Exploratory univariate analyses

Revised threshold for data-defined locations. The large spatial extent of the two data-driven clusters (Fig. 2) indicates that our preregistered threshold may have been too lenient, given the signal-to-noise ratio in the data for the comparison of success-related activation in episodic and semantic tasks versus the control task. We therefore conducted an exploratory analysis to identify smaller data-driven regions within these two clusters, by increasing the statistical threshold to $P < 0.05$, corrected for family-wise error using statistical parametric mapping random field theory for peak-level inference. This resulted in 16 clusters that contained at least 60 voxels (15 additional clusters of fewer than 60 voxels were ignored). These 16 regions were: left and right angular gyri (L_AnG, R_AnG), left and right hippocampi (L_Hipp, R_Hipp), left middle frontal gyrus (L_IFG), two regions within the left and two regions within the right middle temporal gyri (L_MTGa, L_MTGb, R_MTGa, R_MTGb), left precuneus (L_PrC), left superior frontal gyrus (L_SFG), ventral diencephalon (VentDC), right middle frontal gyrus (R_MFG), right precentral gyrus (R_PrG) and two regions within the right cerebellum (R_cereb_a, R_bereb_b). A full description of all clusters is provided in Supplementary Table 1, and image masks are available on the study's OSF page ('Data availability' section).

We extracted trial-based betas from each of the 16 ROIs, as described above for the registered univariate analysis, but fitted the Bayesian linear mixed-model multiple times using different pairs of ROIs each time, until all pairs had been contrasted (120 pairs in total, one for each possible pair of regions). This exploratory analysis therefore included many comparisons and so to avoid an inflated false positive rate, we adjusted the BF criterion for conclusive evidence. To this end,

https://doi.org/10.1038/s41562-025-02390-4

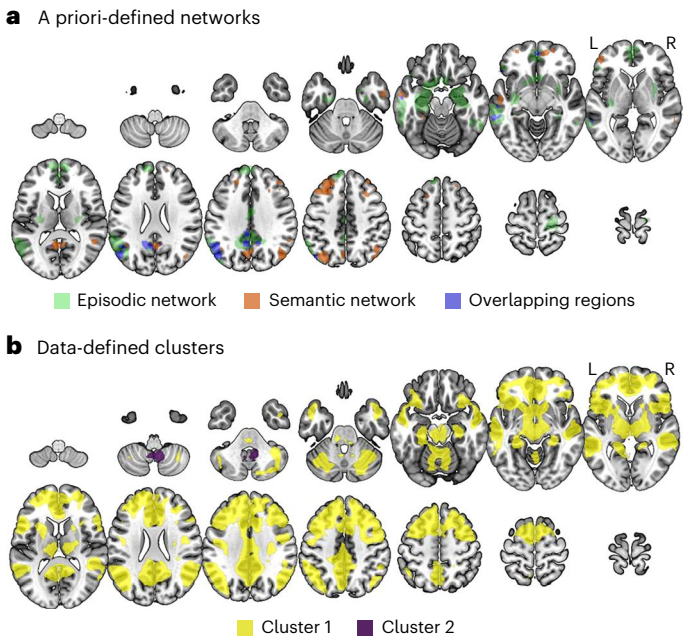


Fig. 2 | Locations of interest. a,**b**, Locations of a priori-defined networks (**a**) and data-defined clusters (**b**). L, left; R, right.

we used simulated data with a null effect to get the false positive rates for one comparison and for 120 comparisons. Based on these results, we repeatedly adjusted the BF criterion until the false positive rate obtained with 120 comparisons was the most similar to that obtained with one comparison. Using this approach, we found that BFs should be adjusted by 0.0222, that is, 'corrected' $BF_{10} = 10/0.0222 = 450.45$.

The data from the 16 data-defined ROIs are shown in Fig. 5a, top, whereas the results of the pairwise comparisons across ROIs are summarized in Fig. 6a, and are fully detailed in Supplementary Table 2. H_0 was conclusively ($BF_{01} > 10$) favoured over the alternative for 41 of 120 comparisons. H_0 was also inconclusively preferred for 49 additional comparisons ($BF_{01} > 1$). For the remaining 30 comparisons, H_1 was preferred ($BF_{10} > 1$), with $BF_{10} > 10$ obtained for 12 of these. However, only two of these survived the 'corrected' threshold of 450.45 described above. Both involved the R_PrG ROI—a region that covers the right precentral gyrus and middle frontal gyrus—which showed a greater episodic than semantic recall success effect; a difference that was greater than in (1) a region within the left middle temporal gyrus (L_MTGb) and (2) a region within the right cerebellum (R_cereb_a), both of which showed the opposite pattern of greater activation for semantic than episodic success. In sum, when using more focal, data-defined ROIs by using a more stringent thresholding the initial task versus control success effect, most comparisons across pairs of these ROIs showed either inconclusive results or conclusive support for H_0, that is, no difference in the size of their episodic versus semantic recall success effects. Nevertheless, we did find two comparisons that showed the predicted crossover interaction in episodic and semantic recall success, even after correcting for the multiple tests performed, thereby providing some limited support for dissociable neural correlates of semantic and episodic retrieval.

Analysis of ATL involvement in semantic versus episodic retrieval. As explained above, one main motivation for using multi-echo fMRI sequence was to better capture mnemonic effects in the ATL. However, in our preregistered analyses, the ATL was included as part of a broader semantic network—and as part of a large data-defined cluster—but not on its own. Therefore, to further examine mnemonic effects within this region, we extracted time series data from the ATL location using a mask that originated from the a priori-defined semantic network of interest but only included the ATL. We then applied the same methods used in all other analyses and fitted the following model: betas ≈ task × response type + (1 + task + response type|participant).

We reasoned that a conclusive interaction between response type and task, such that a greater effect of response type is observed in the semantic versus episodic task, would suggest that the ATL is specifically involved in semantic retrieval. Nevertheless, this analysis showed that the null hypothesis of no difference was conclusively preferred ($BF_{01} = 20.51$), indicating that the ATL is similarly involved in both tasks.

Analysis of 'response type' main effect. To ensure that the relative dearth of three-way interactions between retrieval success versus failure, episodic versus semantic task and brain network does not result from a lack of an overall retrieval success effect, we calculated the BF for the hypothesis that univariate activation in success trials is greater than in failure trials, averaged across episodic and semantic tasks (that is, a hypothesized main effect of the 'response type' factor). This revealed highly conclusive BFs in favour of the alternative hypothesis in a priori-defined networks ($BF_{10} = 2.2 \times 10^{10}$). Note that there is little point repeating this analysis for the data-defined clusters, as these were initially selected to show a stronger success effect in the episodic and semantic tasks than the control task.

Similarity analyses

A priori-defined networks. Similarity values for 'same' and 'different' pairs across tasks and networks are shown in Fig. 4a, right. Logo-brand Fisher-transformed correlation coefficients between 'same' pairs and 'different' pairs were submitted to a Bayesian linear mixed-model described above. Contrary to our prediction, the analysis favoured H_0 over the alternative hypothesis ($BF_{01} = 23.26$).

Data-defined clusters. The same analysis was conducted for data-defined clusters. H_0 was favoured over the alternative ($BF_{01} = 5.88$), but this did not reach our prespecified criteria of $BF > 10$, and therefore, this result is considered inconclusive.

Exploratory similarity analyses

Revised threshold for data-defined locations. We used the same 16 ROIs used in our exploratory univariate analysis to further explore similarity effects, that is, with logo-brand correlation coefficients as the dependent variable for each pair of ROIs (120 comparisons in total). Correlation data from the 16 data-defined ROIs are depicted in Fig. 5b.

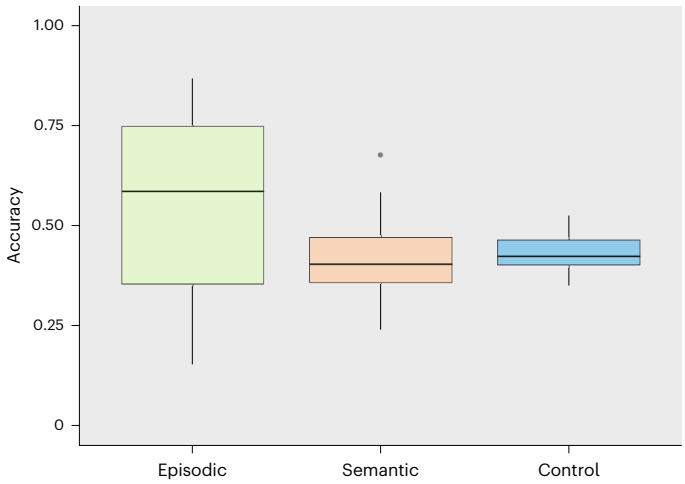


Fig. 3 | Behavioural results from n = 36 participants. The box plots of accuracy rates in the episodic (green), semantic (red) and control (blue) tasks. The horizontal lines represent median values, and the upper and lower hinges correspond to the first and third quartiles, respectively.

https://doi.org/10.1038/s41562-025-02390-4

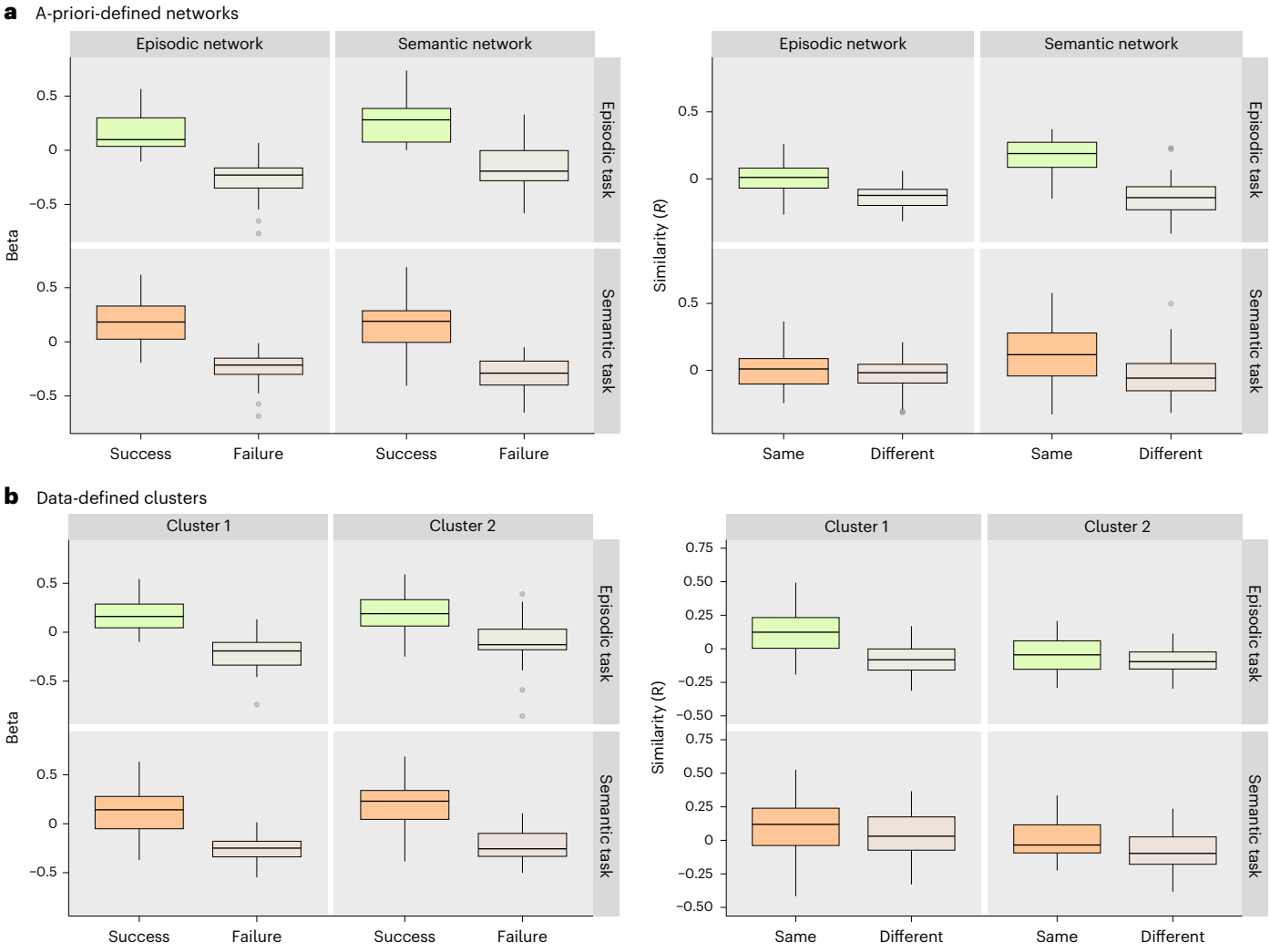


Fig. 4 | Data from n = 36 participants in the two a priori-defined networks and the two data-defined clusters. a,**b**, A priori-defined networks (**a**) and data-defined clusters (**b**). Left: box plots (with outliers as points) of beta values for recall success trials (opaque) and recall failure trials (transparent), used for the univariate analyses. Right: correlation values for 'same' trials (opaque) and 'different' trials (transparent), used for the similarity analyses. In all panels, the data are shown for the episodic task (top, green) and semantic task (bottom, red). The horizontal lines represent median values, and the upper and lower hinges correspond to the first and third quartiles, respectively.

The results of these analyses are shown in Fig. 6b, and are fully detailed in Supplementary Table 3. H_0 was conclusively ($BF_{01} > 10$) favoured over the alternative for 55 out of 120 comparisons and inconclusively preferred for 56 additional comparisons ($BF_{01} > 1$). For the remaining nine comparisons, H_1 was preferred ($BF_{10} > 1$), with conclusive support ($BF_{10} > 10$) obtained for two of these comparisons, but these did not survive the correction to control false positive rate. Thus, when using exploratory data-defined ROIs, most comparisons showed conclusive or inconclusive support for H_0, with some showing inconclusive (and negligible) support for H_1.

Analysis of ATL involvement in semantic versus episodic representation. To explore whether the ATL is specifically involved in semantic representation, we used the same approach described above for the corresponding univariate analysis but using this model instead: coefficient ≈ task × trial type + (1 + task + trial type|participant). Similar to the univariate analysis, the similarity analysis showed conclusive support for the null hypothesis ($BF_{01} = 13.65$), providing evidence against specific involvement of the ATL in semantic versus episodic representation.

Analysis of 'trial type' main effect. As with the univariate analysis, we investigated whether the lack of a three-way interaction resulted from an overall null similarity effect. We calculated the BF for the hypothesis that the similarity between 'same' logo- and word-cue trials will be greater than between 'different' trials (that is, an hypothesized main effect of the 'trial type' factor). This analysis revealed conclusive support for the alternative hypothesis, in a priori-defined networks (BF_{10} = 56). We repeated the analysis for data-defined clusters, as, unlike for the univariate analysis, the selection of the clusters was independent of the similarity effect. This also revealed conclusive support for the alternative hypothesis ($BF_{10} = 2.10 \times 10^6$).

Random selection of 'different' trials. For the similarity analysis, we used all the 'same' pairs that were available for each participant in each task. The correlation between 'same' pairs was compared against the correlation between 'different' pairs, in which each logo-cue trial was matched with one name-cue trial. However, for 'different' pairs, each logo-cue trial might be matched with multiple name-cue trials, as there might be several unassociated trials that were presented in the same block and received the same classification. Previously, for each 'same' trial we selected one matched 'different' trial that was presented closest in time to the associated item. However, to ensure that the obtained (null) results were not due to this '(un)lucky draw', we repeated the

https://doi.org/10.1038/s41562-025-02390-4

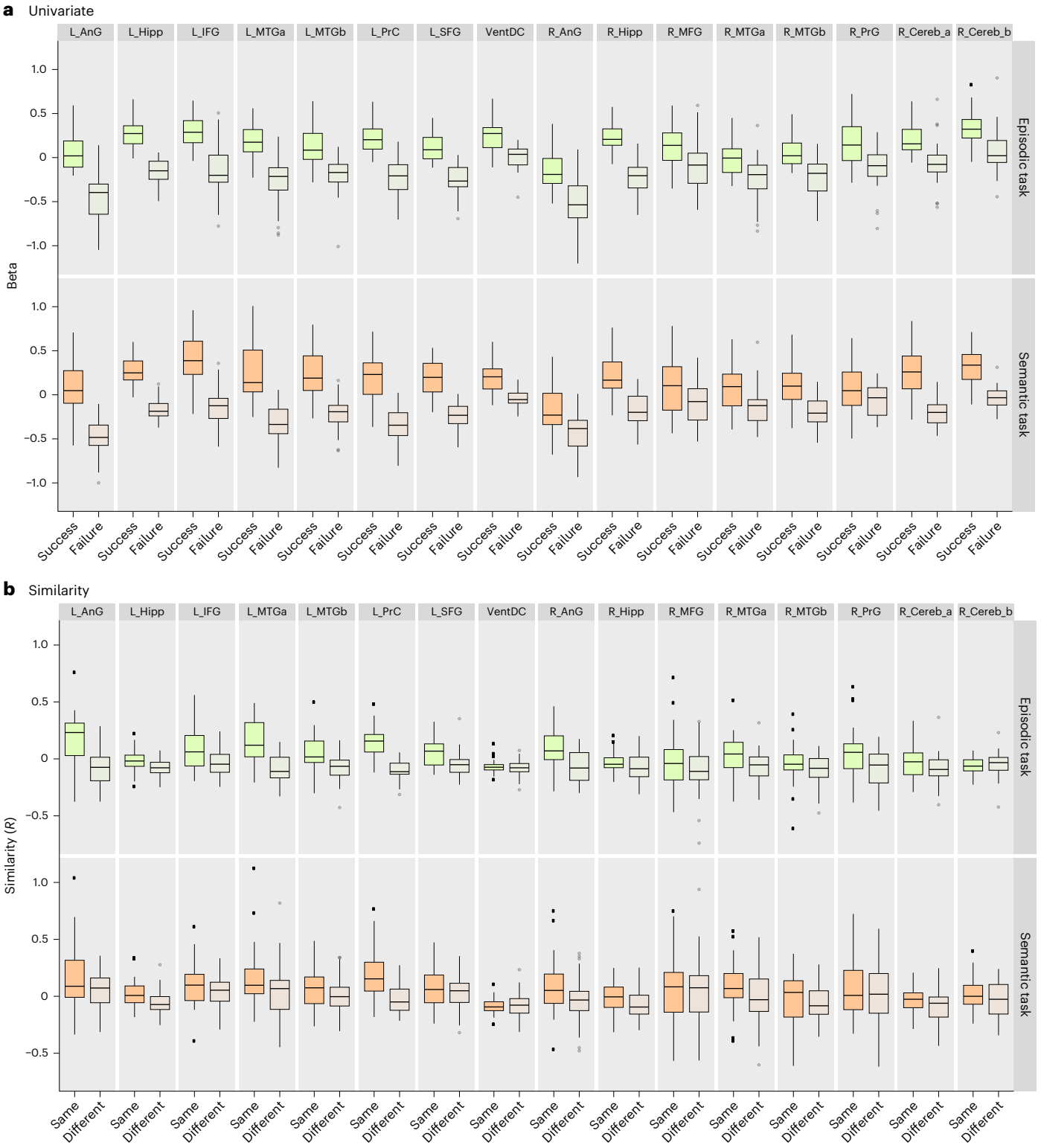


Fig. 5 | Data from exploratory data-defined ROIs with n = 36 participants. **a**, Box plots (with outliers as point) of beta values for recall success trials (opaque) and recall failure trials (transparent), used for the univariate analyses. **b**, Similarity values for 'same' trials (opaque) and 'different' trials (transparent), used for the similarity analyses. In both **a** and **b**, data are shown for the episodic task (top, green) and semantic task (bottom, red), across 16 ROIs. ROIs are left angular gyrus (L_AnG), left hippocampus (L_Hipp), left inferior frontal gyrus (L_IFG), two regions within the left middle temporal gyrus (L_MTGa/b), left precuneus (L_PrC), left superior frontal gyrus (L_SFG), ventral diencephalon (VentDC), right angular gyrus (R_AnG), right hippocampus (R_Hipp), right middle frontal gyrus (R_MFG), two regions within the right middle temporal gyrus (R_MTGa/b), right precentral gyrus, which also includes parts of the middle frontal gyrus (R_PrG), and two regions within the right cerebellum (R_Cereb_a/b). The horizontal lines represent median values, and the upper and lower hinges correspond to the first and third quartiles, respectively.

selection process ten times, such that multiple datasets with various matches between 'same' and 'different' trials were created. The analyses were then repeated for each of these datasets in a priori-defined networks. To reduce computational demands, we ran these models with four chains, 4,000 iterations, and of these, 2,000 were warm-up iterations. We acknowledge the result of this reduction is reduced stability of the BFs that were estimated. The following BF01s were obtained with this analysis: 25.64, 7.14, 15.87, 7.69, 14.71, 30.30, 23.81, 22.73, 8.33 and 29.41. Thus, the null hypothesis was favoured over the alternative in all iterations, and conclusive support was obtained in seven out of ten iterations.

Discussion

In this study, we aimed to distinguish activations and representations during retrieval of semantic versus episodic memories using closely matched tasks. First, we attempted to dissociate episodic and semantic processing by comparing mean activation for successful versus failed retrieval. Much prior work suggests that the brain regions that support successful episodic retrieval are distinct from those that support successful semantic retrieval. Therefore, when contrasting successful versus failed retrieval, we predicted greater overall activation in episodic brain regions for the episodic task compared with the semantic task and vice versa for semantic brain regions. Our analysis of data obtained from a priori-defined episodic and semantic networks conclusively supported the null hypothesis of no difference between episodic and semantic recall in these brain regions. The analysis of data obtained from our preregistered data-defined clusters also favoured the null hypothesis, but evidence was not conclusive, whereas an exploratory analysis performed specifically within the ATL region showed evidence against specific involvement of this region in semantic versus episodic memory retrieval. Nevertheless, when using exploratory data-defined ROIs, following a revision of our preregistered threshold, conclusive evidence for the alternative hypothesis was obtained for 2 out of 120 comparisons. In these comparisons, a greater recall success effect was observed in a region that includes parts of the right precentral gyrus and middle frontal gyrus (R_PrG) in the episodic task, but in regions within the left middle temporal gyrus (L_MTGb) and within the right cerebellum (R_cereb_a) in the semantic task, thereby providing some limited support for dissociable neural correlates of semantic and episodic memories.

We further attempted to dissociate episodic and semantic representations by correlating activation patterns for different cues targeting the same versus different episodic/semantic memories instances. We argued that a multivoxel approach is potentially more sensitive than univariate comparisons of mean activation across voxels, as it might capture more subtle spatial differences between conditions. Therefore, we reasoned, such comparison of multivoxel patterns in these core systems can potentially capture differences in the representations of retrieved episodic versus semantic memories. Regions in which this effect is obtained are assumed to represent episodic/semantic information (though arguably might also support retrieval of that information, at least to some extent[29]). Prior theorizing suggests that the brain regions that hold episodic representations are distinct from those that hold semantic ones, and we therefore predicted greater similarity in episodic brain regions for the episodic task and greater similarity in semantic brain regions for the semantic task. Our analysis of data obtained from a priori-defined networks revealed conclusive support for the null hypothesis, and analysis of data obtained from data-driven clusters showed inconclusive support for the null. When using exploratory data-defined ROIs, evidence for the null hypothesis of no difference was obtained for most comparisons.

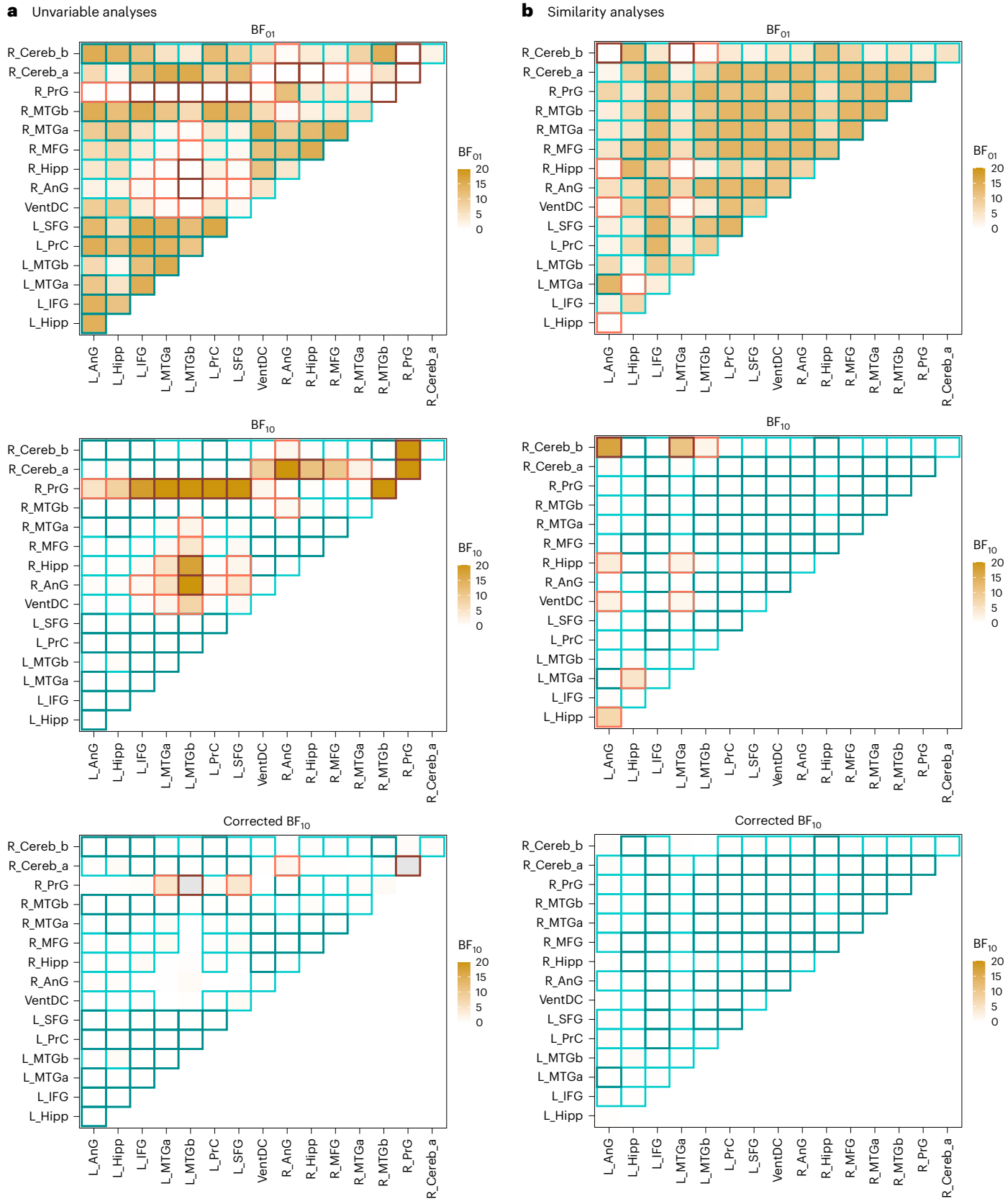
Taken together, our study shows very little evidence for distinct processing of and no evidence for (or evidence against) distinct representation of semantic and episodic memories. Although these results are aligned with other functional neuroimaging studies of healthy participants[10], in which considerable overlap between brain regions involved in episodic and semantic processing was observed, they stand in contrast to decades-long evidence from neuropsychological studies, repeatedly supporting the proposed distinction. Importantly, neuropsychological studies do not only propose a distinction between episodic and semantic memories but also pinpoint this distinction to specific brain regions, namely the MTL and the ATL, believed to support episodic and semantic memory, respectively. This dissociation was not supported by the current neuroimaging study, despite what we believe is the most closely controlled comparison yet performed.

A region that did show more activation for episodic versus semantic recall is the right PrG/MFG. This finding converges with previous studies showing activation in this region during episodic retrieval[30–33]. Notably, these previous studies further suggest that this activation is more likely to reflect strategic processes rather than retrieval per se. Indeed, activation in this region is also frequent during visual attention tasks, suggesting that it might play a role in attentional or other overarching processes rather than purely mnemonic ones[30]. One established suggestion is that this region is involved in retrieval monitoring, that is, the evaluation of the products of memory retrieval with respect to their relevance to the retrieval task[31], and the utilization of that information to guide subsequent behaviour[32]. In a previous study[33], the amount of retrieval monitoring was manipulated directly by contrasting recognition for words in a task that requires reference to the spatiotemporal context of words presented during a previous study episode ('exclusion' condition) with recognition of words in a task that does not require such reference ('inclusion' condition). The authors reasoned that old words in the exclusion condition that were studied in the inappropriate context require greater monitoring, because successful recollection of the study context is necessary to overcome the sense of familiarity associated with old words. The study showed that this condition, in which greater monitoring requirements were imposed, was associated with greater activation in the right MFG. One possibility is that in the current study, the episodic task required additional monitoring compared to the semantic task. For example, it could be that some familiarity might have been experienced for semantic associations of logos in the episodic task, which therefore required overcoming potential semantic interference to successfully recall the episodic information, arguably posing greater monitoring requirements.

Fig. 6 | Results of univariate and similarity exploratory analyses, conducted for 16 data-defined ROIs. a, Univariate exploratory analyses. **b**, Similarity exploratory analyses. Each cell of the matrices represents the BF obtained for the three-way interaction (ROI × task × response type (for the univariate analyses); ROI × task × trial type (for the similarity analysis)), for each pair of ROIs. Results are shown for BF_{01} (top), for BF_{10} (that is, the inverse of BF_{01}; middle) and for BF_{10} after a correction for multiple comparisons was applied (see text; bottom). The colour scale, ranging from white to brown, indicates the value of the BFs. Note that values were capped at 20 to aid visualization. See Supplementary Tables 2 and 3 for precise numbers. Comparisons that yielded conclusive support for H_0 ($BF_{01} > 10$) are highlighted with dark cyan, and those that yielded inconclusive support for H_0 ($BF_{01} > 1$) are highlighted with light cyan. Similarly, comparisons that yielded conclusive support for H_1 ($BF_{10} > 10$) are highlighted with dark coral, and those that yielded inconclusive support for H_1 ($BF_{10} > 1$) are highlighted with light coral. ROIs are left angular gyrus (L_AnG), left hippocampus (L_Hipp), left inferior frontal gyrus (L_IFG), two regions within the left middle temporal gyrus (L_MTGa/b), left precuneus (L_PrC), left superior frontal gyrus (L_SFG), ventral diencephalon (VentDC), right angular gyrus (R_AnG), right hippocampus (R_Hipp), right middle frontal gyrus (R_MFG), two regions within the right middle temporal gyrus (R_MTGa/b), right precentral gyrus which also includes parts of the middle frontal gyrus (R_PrG) and two regions within the right cerebellum (R_Cereb_a/b).

https://doi.org/10.1038/s41562-025-02390-4

Other regions involved in the abovementioned interactions, which showed increased activation for semantic versus episodic recall, are the left MTG and a region within the right cerebellum. The MTG has been shown before to be involved in a range of tasks that require semantic processing. For example, increased activation in this region was observed when conceptual-semantic processing (analysis of meaning) was contrasted against lexical and phonological processing[34], in the presence of semantic overlap between consecutive items (that is, a semantic enhancement effect)[35], when processing person-related semantic knowledge of famous faces[36] and in tasks that pose high



versus low semantic demands[37]. The right cerebellum was also implicated in semantic processing in previous studies. In particular, the cerebellum was shown to be involved in semantic processing during reading[38–40] and, recently, to also be causally involved in semantic memory[41]. Despite these compelling links with previous literature, it is important to note that without sufficient evidence for a crossover interaction, let alone a reversed association[42], in which one region selectively prefers episodic processing and another semantic processing, any strong conclusions regarding the dissociation between these processes is unwarranted. Given the exploratory nature of the current results, our overall conclusion is that, with the current study, we are unable to provide conclusive support for the distinction between episodic and semantic recall in healthy adults.

What might underlie this difficulty to observe the predicted distinction between episodic and semantic memories, which is indeed well established in the neuropsychological literature? One possibility is that in the healthy brain, both 'episodic' and 'semantic' locations are involved when information is recalled, be it episodic or semantic, even though the involvement of 'semantic' regions during episodic recall (or vice versa) is not essential. Although this notion is likely to be an accurate description of recall in everyday life, it is unlikely that it fully accounts for the overlap that was observed with the experimental paradigm used in our study. As explained in the Main, although episodic information might scaffold or facilitate semantic retrieval and vice versa, our design meant that such facilitation would not determine success unless some initial information from the probed process (that is, episodic information in the episodic task or semantic information in the semantic task) was obtained. It is possible, however, that even though the design does produce cognitive differences, the neural differences associated with them are not observable. For example, it might be that the neural differences associated with successful recall of episodic versus semantic memories are much more subtle than those associated with their facilitation or scaffolding and are therefore masked by the latter. If this is the case, the predicted differences might become observable when a similar paradigm is used with different neuroimaging methods or protocols. For example, the proposed 'masking' effect might be less pronounced if the temporal profiles of the retrieval processes are considered, for example, it might be that the temporal profile associated with successful recall is short-lived, whereas the profile associated with mnemonic scaffolding is more sustained. Therefore, differences between episodic and semantic processing might be more easily observed with methods that provide greater temporal resolution than fMRI, such as electroencephalogram or magnetoencephalography.

Another alternative is that episodic and semantic recall are only distinct under certain conditions or in certain populations. For example, differences between healthy adults and patients might result from some level of reorganization or compensation that affects processing. Hence, it could be that a particular brain region (for example, the MTL) is 'in charge' of both semantic and episodic recall, but when the region becomes dysfunctional, other regions (for example, the ATL) are able to instead support some but not all of its functionality (in this case, semantic but not episodic memory). Although this possibility cannot be ruled out entirely, it was shown that, at least for Alzheimer's disease, functional compensation increases activity within networks that are associated with task performance in healthy individuals (quantitative change in brain activation), rather than involves recruitment of brain regions that are normally unassociated with the task[43]. If this is the case, we would still expect to observe a dissociation in the healthy population, though possibly to a lesser degree. Another factor that might contribute to the discrepancy between previous neuropsychological studies and the current study is the characteristics of the current sample. In particular, in neuropsychological studies patients with Alzheimer's disease and semantic dementia are often older compared with the younger healthy participants who were tested here. Possibly, longitudinal changes in the segregation of semantic and episodic memories mean that a dissociation can be observed in older but not younger adults. However, we believe that this explanation is unlikely for two reasons. First, in most neuropsychological studies, patients are compared with age-matched controls. Therefore, if increased segregation is solely due to ageing, then it should not be observed in neuropsychological studies either. Second, this notion implies that the episodic and semantic systems overlap during young adulthood and become more segregated with cognitive ageing. Nevertheless, previous research shows that brain networks are less functionally distinct in older versus younger adults during both task and rest[24,44,45], suggesting that the opposite pattern is more likely: as people grow older their brain networks and cognitive functions become less rather than more segregated.

Our above attempts to explain the current null results all suggest that the distinction between semantic and episodic retrieval does exist, but might occur (or might be detected) under specific conditions. Nevertheless, an important possibility to consider is that episodic and semantic memories indeed involve the same neural mechanisms. This view might apply specifically to the retrieval phase, or it might broadly encompass declarative memory. If the former, then the dissociations observed in neuropsychological studies might be attributed to distinct encoding or retention of episodic and semantic memories. For example, it could be that patients with Alzheimer's disease are less able to encode novel associations compared with healthy controls. This reduced ability to encode information will then also lead to reduced ability to retrieve information, but not because of difficulties in retrieval per se, but because less to-be-retrieved information is available. Notably, however, in neuropsychological studies, it is often impossible to distinguish between different mnemonic stages, as data preceding the lesion/disease is usually unavailable for patients. Neuroimaging studies can more easily address this question of different neural mechanisms during different mnemonic stages, although unfortunately, this distinction cannot be established in the current study. Namely, we only collected neuroimaging data during retrieval, and indeed, obtaining encoding data would have been impossible with the current design in which semantic knowledge is assumed to be attained over prolonged time periods. Given the current results, and the observed discrepancy between patients and neuroimaging studies, the field would benefit from future studies that are designed to explore this proposal, that is, that whereas encoding varies for episodic and semantic memories, their retrieval involves overlapping neural mechanisms.

As noted, however, a final possibility to consider is that the distinction between semantic and episodic memories is in fact outdated, or more nuanced than previously thought. As mentioned in the Main, previous neuroimaging studies have already highlighted more similarities than differences when episodic and semantic memories were compared within the same design[16–22]. Recent theoretical perspectives and a growing body of empirical evidence further suggest alternative approaches to the episodic–semantic distinction. These approaches can be broadly referred to as continuum approaches, which view episodic and semantic memories as opposite ends of a continuum, and multidimensional approaches, which position memories in a multidimensional space, allowing them to vary across a range of categorical or continuous dimensions[10,46,47]. A recent example for the latter is the multidimensional model of mental representations, which suggests that mental representations vary along continuous dimensions in how much they are temporally specific or general, perceptual or conceptual, and idiosyncratic or shared[47]. As such, this model arguably teases apart dimensions that are often confounded when the episodic–semantic distinction is considered. Two recent studies support such a refined view of episodic and semantic memories. In an fMRI study, brain activity was recorded while participants verified statements concerning general (semantic) facts, autobiographical facts, repeated events and unique

(episodic) events. Semantic and episodic memories involved shared neural correlates, but the degree of engagement of each region within this network varied across different memory types[48]. In another study, electrical brain stimulation was used to trigger a transient brain state in epileptic patients while neural activity was recorded by intracranial electroencephalogram. Such triggers often evoke sudden involuntary reminiscences that can vary in their degree of contextual dependency. The study provided some suggestive evidence of greater functional connectivity for more complex memories, with a decreasing gradient of complexity from episodic memories, to personal semantics and to familiar memories[49]. Similar to our current findings, these findings resonate better with theories that assume a continuum between different memory systems or with a multidimensional perspective of declarative memory where different memory types vary in magnitude of activation within a common network of brain regions, than with the traditional view that separates the systems apart.

It is worth noting that in many neuropsychological studies (as in most neuroimaging studies), the episodic and semantic tasks used to investigate their dissociation were not always closely matched. For example, in a study that showed profound episodic memory deficits but relatively intact semantic memory in patients with Alzheimer's disease, the episodic measures included the logical memory subset from the Wechsler memory scale, a face-recognition memory test and the Rey complex figure test. Semantic measures, on the other hand, included a category fluency test, naming of line-drawn objects and response to a verbal description, semantic feature questions, picture sorting and word-picture naming[4]. Although the episodic and semantic tasks used in that study clearly differ in the cognitive processes that they are tapping, they also differ in their procedures, their measures, their administration and potentially other aspects that are not directly related to the type of memory that they involve. Taken together with our current results, we suggest that although the distinction between the cognitive profiles of patients with Alzheimer's disease and semantic dementia is well established in the neuropsychological literature[3–5,7–9,50,51], its mapping into the conceptual distinction between episodic and semantic memory would benefit from further investigation using approaches that allow more direct comparison.

Before concluding, some caveats and limitations of the current study should be considered. First, despite our efforts to match the episodic and semantic tasks as much as possible, some differences remain. Specifically, while semantic associates were acquired in natural settings and over an extended time period, episodic associates were acquired in lab settings and over a limited time period (that is, during the experimental session). These potentially confounding differences might have been able to offer an alternative account for a distinction, if observed. Nevertheless, as the current study produced mainly null results, this potential confound does not undermine our key conclusion. If the task is to be used in future studies (for example, with magnetoencephalography, as suggested above), further optimization to avoid this issue should be considered. Furthermore, it could be that with our well-matched controlled tasks, we forced participants to engage in cognitive processes that are not natural and potentially differ from how semantic and episodic recall are achieved in 'real life' circumstances. Although this potential limitation should not be taken lightly, it can equally apply to most lab-based experiments and indeed inspires a recent shift towards studies of greater ecological validity. However, studies such as the current one, in which the main aim is to disentangle cognitive processes that are highly intertwined in natural circumstances, inevitably require some level of artificiality to permit sufficient control.

Another potential limitation is that our selection of data-driven locations was tuned to detect univariate effects: using a whole-brain univariate analysis, we selected regions in which the success effect in the two critical tasks (episodic and semantic) was greater than in the control task. Nevertheless, the same locations were also used for the similarity analyses, although for these analyses, they might have not been ideal. Theoretically, it is possible that episodic and semantic processing and representation involve distinct regions, with our analysis optimized to detect the former but not the latter. Notably, due to the structure of the control task which only included one type of subblock, whole-brain similarity analysis that facilitates the control task was not possible, as this analysis requires the computation of similarity between probes across two subblocks. Nevertheless, the possibility remains that we were not able to detect any similarity effects with the current study due to this suboptimal selection of locations.

An additional limitation concerns the fMRI protocol selected for the study. Namely, while the multi-echo protocol used here crucially allowed us to recover ATL signal, it might have reduced our ability to detect hippocampal activation. A recent study compared the ability of different fMRI protocols to detect activation during a semantic task[52]. In their study, the number of echoes (single-echo time (TE 30 ms) versus multi-echo (TE 12, 25.85 and 38.70 ms)) and bands (single-band, multiband (2)) were manipulated independently to construct a 2 × 2 factorial design in which the effects of multi-echoes and multibands can be estimated. A whole-brain analysis revealed greater magnitude and precision of activation in the hippocampus when single- versus multi-echo protocols were used. Nevertheless, a study that examined hippocampal activation during an episodic memory task showed clear hippocampal involvement in the task when multi-echo data were used (TE 12.5, 27.6 and 42.7 ms) together with Multi-Echo Independent Component Analysis (ME-ICA), but not when single-echo data (TE 26.6 ms, albeit acquired within the same multi-echo protocol) were used[53]. This suggests that while the possibility of some reduced ability to detect hippocampal activation remains, the application of ME-ICA in our study was potentially helpful in balancing trade-offs between the ability to detect ATL versus hippocampal signals.

Finally, a potential limitation regards the statistical approach that was used to correct for multiple comparisons in our exploratory analyses. With Bayesian statistics, there is currently no wide agreement as to what approach (if at all) should be taken to reduce false positive rates due to multiple comparisons, and it is plausible that the correction that we have applied here was too strict. Further examination of the comparisons where BFs were conclusive before a correction was applied (that is, uncorrected $BF_{10} > 10$) showed that most of these involved the right PrG/MFG, in which a greater recall success effect is observed in the episodic task, and an additional 'semantic' region (L_MTGa/b, R_MTGb, L_IFG, L_SFG, R_cereb_a/b). As mentioned above the PrG/MFG often reflects retrieval monitoring rather than retrieval per se, and therefore, it is unclear whether these comparisons can provide strong evidence for the alternative hypothesis. Two additional comparisons, for which the uncorrected BF exceeded our predefined threshold revealed a greater recall success effect in the R_AnG and R_Hipp for episodic recall, but in L_MTGb and R_cereb_a for semantic recall. Notably, these two 'episodic' regions are closer to what one might expect from an episodic–semantic dissociation. Specifically, studies of neurological patients and functional neuroimaging in humans have implicated the hippocampus as a critical region for the formation and retrieval of episodic memories[5,54–57]. As for the angular gyrus, in the cluster that emerged in the current exploratory analysis, recall success effects in the critical versus control tasks covered locations within the right posterior AnG, typically observed in episodic recall tasks[30,58–60], and particularly when episodic events are remembered vividly[61]. Of note, the cluster extended further into the right anterior AnG, into locations that are not typically associated with episodic recall. We explored further whether the greater recall success effect for episodic versus semantic memory in the right AnG does not reflect the different nature of the tasks (episodic versus semantic) but instead might be due to variations in vividness between the tasks. Although vividness ratings were not collected in the current study, a reasonable proxy is participants' reports of whether they were able to retrieve few or many

associated details for each probe. An exploratory analysis did not show conclusive differences in 'detailedness' for episodic versus semantic memory. Thus, taken together, the possibility remains that the right hippocampus/AnG and the left MTG/right cerebellum are candidate ROIs in which the episodic–semantic distinction indeed exists and should be explored further in future research.

To conclude, despite employing tasks that are more closely matched than typically achieved, we were unable to offer decisive conclusions regarding the dissociability of episodic and semantic memories during their recall. These results, which are hard to reconcile with decades-long evidence from neuropsychological studies, call for further research using alternative approaches, as well as for further specification of the components of declarative memory in terms of their association with patients' profiles and their neural underpinning at different mnemonic stages.

Methods

The methods used in this study carefully followed those that were registered with our Stage 1 report. Any deviations are summarized in the 'Deviations from registered protocol' section.

Ethics information

This study was approved by a local ethics committee (Cambridge Psychological Research Ethics Committee reference PRE.2020.018). Participants provided informed consent and were compensated for their time with £50.

Participants

The sample size for the study was determined using a Bayesian 'sequential design with maximal n'[62], where the maximal feasible number of participants, given time and funding, was $n = 100$. Using this approach, the experiment was due to run in batches with 40 participants in the first batch and 10 participants in each additional batch, calculating the BF in favour of our main hypotheses. Following simulations (fully detailed in Supplementary Analysis 1), the stopping criteria was set to be $n = 100$, or $BF_{10} > 10$, or $BF_{01} < 1/10$ for either of the two main analyses (univariate or similarity) for the a priori-defined networks of interest.

Forty participants were recruited from the MRC Cognition and Brain Sciences' SONA system or from word-to-mouth. This initial batch was sufficient for the BF for the null to exceed our stopping criterion (that is, $BF_{10} < 1/10$) for both the univariate and similarity analyses (Results). Data exclusion followed the criteria in our Stage 1 report, which specifies that participants who are excluded before the analysis of their data will be replaced with others. According to the report, reasons for preanalysis exclusion were failure to complete the task, failure to arrive to the second session, noncompliance with inclusion criteria, faulty equipment and experimenter error. Accordingly, five participants were excluded before the analysis of their data—one due to an incidental magnetic resonance imaging (MRI) finding, two due to technical issues, one due to experimenter error in loading the sessions and one who did not show up for the second session—and were replaced with others. Our Stage 1 report further indicates that participants who were excluded following the analysis of their data will be removed from the relevant analyses but will not be replaced. Potential reasons for postanalysis exclusion were lack of trials in one or more experimental conditions (due to floor/ceiling performance), increased movements (two s.d. or more than the mean motion across all participants) and/or motion that resulted in reduced coverage of key locations (within a priori-defined networks, as detailed below). Accordingly, four participants were excluded following the analysis of their data but were not replaced. These included one exclusion due to lack of trials in one of the experimental conditions, one due to increased head movements and two exclusion of two participants for which the 'tedana' denoising algorithm (detailed below) failed to converge during preprocessing.

Our final sample ($N = 36$) included female (25) and male (11) adults (mean age of 24.1 years, s.d. of 5.12, range 18–35), who were native English speakers, magnetic resonance-compatible, with normal or corrected-to-normal vision and not diagnosed with attention deficit hyperactivity disorder, dyslexia or any other developmental or learning disabilities. Participants were pseudo-randomly assigned into the various counterbalancing conditions, to keep the mean age and gender distribution across the various experimental lists roughly equated.

Ceiling/floor performance

Following behavioural pilots (see below), we expected participants to reach ~50% accuracy in each task. Nevertheless, we took precautions to ensure that performance remained interpretable and is not at floor or at ceiling when participants performed the task in the scanner. In our Stage 1 report, we indicated that performance will be assessed after data are collected from the first five participants, and the task(s) will be revised if floor/ceiling effects are detected. Accordingly, performance was assessed after data were collected from the first five participants (one for each of the five stimulus lists). Floor performance was defined as when the average of successful trials was <10%, and ceiling performance were defined as when the mean of successful trials was >90% in any of the tasks. Assessment of the data indicated that there were no floor/ceiling effects in any of the tasks: in the episodic task, mean accuracy across the first five participants was 36% (s.d. of 16%, range 15–58%), in the semantic task, mean accuracy was 45% (s.d. of 15%, range 24–68%), and in the control task, mean accuracy was 41% (s.d. of 0.044%, range 35–46%).

Materials

The entire stimulus database is available at https://osf.io/dm47y/. Stimuli for the episodic and semantic tasks were names and logos of brands from various categories (technology, food and drinks, travel, entertainment, clothing, finance, organizations, sports, superheroes and household). The initial stimulus pool for the episodic and the semantic tasks included 440 logos obtained from various internet sources. Familiarity of the logos was assessed in a pilot study with eight participants (all native English speakers, aged 25–39 years, five females). Participants viewed the logos and indicated whether and to what degree they were familiar with them. Each trial of the pilot study started with a 1 s fixation cross, followed by the presentation of the logo for a duration of 1 s with the familiarity scale presented below the image ('1, not familiar; 2, somewhat familiar; 3, highly familiar; 4, can name'). The logo then disappeared, but the scale remained on the screen until the participant provided their response. They were instructed to use the 'can name' option only if they can name the actual brand associated with the logo, and selection of this option prompted another screen in which they had to type in the brands' name. They were instructed to avoid guessing and 'gist-based' responses (for example, responding 'car' instead of 'Toyota').

Based on the data from the pilot study, we constructed a stimulus database with 240 entries. Each entry included a logo and its associated brand name. This database contained 80 logos that were highly nameable (were named correctly by five or more of the participants in the pilot study), 80 that were somewhat nameable (were named correctly by one to four participants) and 80 that were not nameable. A total of 32 additional entries were selected to be used as fillers, examples and practice trials. The stimulus database was then divided into 10 subsets of 24 entries. Subsets were matched in terms of stimulus familiarity (1–4 familiarity rating provided by the participants) and stimulus category such that each list included an equal number of stimuli from the abovementioned categories. Within each subset, entries were randomly ordered and then associated with the following one (and the last entry within each subset was associated with the first one), to add to this entry an unrelated brand's name. Thus, each entry included three elements: the logo, the associated brand's name and an unrelated alternative. The experimental lists were then constructed such that

for each participant, different entries were presented for the episodic task, which included 144 entries, and the semantic task, which included 96 entries. For example, a participant that was allocated subsets 1–6 (each with 24 entries, for a total of 144 entries) for the episodic task, was allocated subsets 7–10 for the semantic task. Experimental lists were counterbalanced across participants.

Stimuli for the control task were pseudo-random combinations of two English consonants with added noise. To construct these stimuli, we created 192 random combinations of consonants, from which we chose 80 unique combinations, excluding combinations that convey a well-known semantic meaning (for example, 'EG'). The selection of the stimuli was verified by three independent judges. We then created displays of these combinations. Each display included one combination, shown in a random font from the list of available fonts in Matlab, a random colour and a fixed font size (150). Noise was then added to the displays using the 'impulsenoise' Matlab function[63] at a random level of 85–100%.

Design and procedure

The paradigm used in the experiment is illustrated in Fig. 1. The experiment was administered using E-Prime 2 (Psychology Software Tools). It consisted of two critical tasks, the episodic task and the semantic task, which took place on two different days, and an additional control task which was completed together with the semantic task. The order of the critical tasks was counterbalanced across participants. In all tasks, participants were instructed not to guess and to only provide a verbal response if they were sure about their answer or to say 'pass' otherwise. Verbal responses were recorded using an fMRI compatible noise-cancelling microphone (OptoAcoustics Ltd). All stimuli were presented against a grey background. Pictorial stimuli (logos) were presented at ~6–8 cm in on-screen size, and textual stimuli (brands' names) were displayed in black in Courier New 24-point font.

On arrival at the laboratory, participants were seated in a quiet room, where they signed an informed consent form and completed the unscanned part of the relevant task. Once completed, they were given the instructions and practice trials for the scanned part of the task. Moreover, they were provided with instructions for the MRI scan and were told to try to minimize head movements even though they need to give verbal responses in the scanner. They then performed the relevant task(s) in the MRI scanner. In all tasks and phases, two filler trials were provided before the beginning of each block. These trials were not included in the analyses.

Episodic task. The task included three phases: study phase, a scanned test phase and debriefing. At the beginning of the study phase, participants were told that they will be presented with pairs of unrelated stimuli (a brand's name and a logo) and were instructed to associate them for a subsequent memory test by creating a story such as a situation that could have happened that links the stimuli together (for example, 'Yesterday I was searching for NASA on GOOGLE so I can teach my brother about space rockets'). They then completed a practice block of ten trials, with the experimenter ascertaining that they understood the nature of the associations that are to be generated for the stimulus pairs. The study phase then started. An exemplary study trial is depicted in Supplementary Fig. 1. Each study trial began with a 1 s fixation cross, followed by a 5 s presentation of the stimulus pairs and a 5 s blank screen. Participants were instructed to use this time (10 s in total: stimuli presentation + blank screen) to generate their association. Next, a screen with the text 'easy/hard?' was shown for 2 s, and participants were asked to press one key if they found it easy to come up with an association or another key if they found it difficult. The right index and middle fingers were used for these keypresses and finger assignment was counterbalanced across participants. The study phase was divided into three blocks of 48 stimulus pairs, for a total of 144 stimulus pairs, presented in a random order. A short self-paced break was given between blocks. To obtain adequate subsequent memory performance (see 'Pilots' section), the study phase was repeated twice with stimulus pairs presented in different order, and participants were instructed to think about the same scenario as before, to the best of their ability.

Following the completion of the study phase, participants were given instructions for the test phase, which included logo-cue trials and name-cue trials. For logo-cue trials, participants viewed studied logos and indicated what the associated brand was (that is, the episodic associate, from the study phase) and how many details they remember from the associated story that they had created. Similarly, for name-cue trials, they viewed brand names and indicated how many details they remember from the study phase. Following the completion of a practice block of ten trials and provision of additional instructions regarding the scanning procedure (see above), participants completed the test phase while in the scanner. The procedure used for the test phase was similar to that used in our previous work[61] and was shown to be useful for a memory task that relies on verbal responses. Each logo-cue trial started with a jittered fixation cross (1–8 s; mean of 4), followed by the presentation of the logo for 1 s and a 5 s blank screen. Participants were instructed to use this time to think about the details of the story that they created when associating the stimuli together during the study phase. Next, a screen with the text 'pair associate?' was shown for 3 s. During this time, participants had to verbally recall the name of the associated brand from the study phase. This was followed by a 2 s screen with the text 'many/few?', for which they pressed one key if they managed to retrieve many details from the study phase or another key if they managed to retrieve few (or no) details. The right index and middle fingers were used for these keypresses, and finger assignment was counterbalanced across participants. Key assignment corresponded to the study phase, such that the same finger was used for 'easy' and 'many' and another was used for 'hard' and 'few'. After completing the subblock of 48 logo-cue trials, a subblock of 48 name-cue trials ensued, corresponding to the same stimuli. In each name-cue trial, after a jittered fixation (1–2 s; mean of 1.5), a brand name was presented for 1 s, followed by a 5 s blank screen. Again, during this time, participants were instructed to think about the details of the story from the study phase. This was followed by a 2 s 'few/many?' response screen. The test phase was comprised of three blocks (runs). In each block, a subblock of 48 logo-test trials was displayed, followed by a subblock of 48 brand-test trials. A short break, during which a new run was initiated, was given between blocks.

Participants then performed the final debriefing phase, whose aim was to allow us to control, as best as possible, for semantic processing in the episodic task ('Trial classification' section). In this phase, participants viewed the logos that were presented throughout the episodic task and indicated whether, and to what degree, they were familiar with them before the experiment. Each debriefing trial started with a 1 s fixation cross, followed by the presentation of the logo for a duration of 1 s with the familiarity scale presented below ('1, not familiar; 2, somewhat familiar; 3, highly familiar; 4, can name'). The logo then disappeared, but the scale remained on the screen until the participant provided their response. Selection of the 'can name' option prompted another screen in which participants typed in the brands' name. Data from the debriefing phase was used to inform our analyses. Namely, in the episodic task, participants are likely to know the original pairing for some logos, which might cause semantic interference when the new (episodic) pairings are formed. To minimize this potential confound, trials that were classified as 'can name' during the debriefing session were initially excluded from the analyses but were then reinstated as they conformed to the same numerical trends ('Data analysis' section). To allow sufficient statistical power after the potential exclusion of these trials, the episodic task included more trials than the semantic task (144 versus 96 trials), with the number of trials-per-block (48) remaining constant across tasks.

Semantic task. The task included two phases: a familiarization phase and a scanned test phase. The purpose of the familiarization phase was to expose participants to the stimuli that were presented during the test phase. At the beginning of the familiarization phase, participants were told that they will have to provide two types of judgement: one for words and one for pictures. Exemplary familiarization trials are depicted in Supplementary Fig. 1. For pictures, they viewed logos presented above a bisecting red line and indicated whether the line is presented more to the left or to the right of the picture. We chose this task because it requires participants to attend to the appearance of the stimulus, shown on each trial, but does not involve an explicit identification of the logo or the brands associated with them (see similar justification by others[64,65]). For words, participants viewed brand names and indicated whether their first and last letters are in alphabetical order. This task was chosen because it is challenging, yet does not require semantic processing. Following these instructions and the completion of a ten trials practice block for each judgement, participants completed the familiarization phase. For pictures, each trial started with a fixation cross displayed for 1 s, followed by the presentation of the logo for 3 s. A vertical line bisected the picture, leaving 65% of the picture either on the left or right side of the line. This line was randomly tilted between 10° and 40° from the vertical. After the disappearance of the stimuli, participants indicated by a key press on each trial whether more of the logo was presented to the left or to the right of the line by responding to a screen with the text 'left/right?', shown for 2 s. For words, each trial started with a fixation cross (1 s), followed by the presentation of the brands' name (3 s), and a screen with the text 'alphabetical/non-alphabetical?' (2 s). Participants pressed a key with their index finger if the first and last letters were in alphabetical order and another key with their middle finger if not. The familiarization phase was divided into two blocks. In each block, a subblock of 48 pictures was displayed, followed by a subblock of 48 words, for a total of 96 pictures and 96 words. A short break (self-paced) was given between blocks.

Following the completion of the study phase, participants were given instructions for the scanned test phase. The test phase resembled that of the episodic task, but instead of recalling the associated brand from the study phase, participants were asked to recall the name of the actual (semantic) associated brand from their prior knowledge. Similarly, they were asked to indicate whether they were able to retrieve few/many details associated with the logo/brand, rather than with the associating story. The test phase included two blocks (runs) of 48 logo-cue trials and 48 name-cue trials.

Control task. Because the semantic task has less trials than the episodic task ('Trial classification' section), it takes less time to complete. Therefore, the control task took place on the same day as the semantic task and followed it. The control task was used to control for general success effects that are unrelated to semantic or episodic memory. Therefore, we designed a challenging visual perception task that can produce a roughly equal number of success and failure trials. The instructions for the control task and a block of five practice trials were provided before the scan, after the instructions for the test phase of the semantic task were given. Participants were instructed to identify two-letter combinations of English consonants with added noise. Each trial started with a jittered fixation cross (1–8 s), followed by a letters-display presented for 1 s, and a 5 s blank screen, to match the trial structure of the episodic and semantic tasks. Next, when a screen with the text 'letters?' was shown (for 3 s), participants had to say what letters were presented in the display. Finally, a screen with the text 'many/few' appeared (for 2 s), and participants indicated whether they were able to extract many/few details (colour, shape and so on) while the letters were presented. The control task included two blocks (runs) of 40 trials.

Pilots. Six participants completed behavioural pilot versions of the above tasks. The aim of these pilots was to assure that (1) the instructions are clear; (2) there are no bugs/typos/problems with the tasks; (3) all tasks are feasible and performance are not at floor/ceiling; and (4) the tasks are roughly matched for difficulty. Pilot data, as well as the full list of changes to the tasks that were made during/following these pilots are available at https://osf.io/dm47y/.

fMRI acquisition

The same acquisition and preprocessing protocol were used for all tasks. MRI data were collected using a Siemens 3T PRISMA system with a 32-channel head-coil. Structural images were acquired with a T1-weighted three-dimensional magnetization prepared rapid gradient-echo sequence (repetition time (TR) of 2,250 ms; echo time (TE) of 3.02 ms; inversion time (TI) of 900 ms; 230 Hz per pixel; flip angle of 9°; field of view (FOV) 256 × 256 × 192 mm; GRAPPA acceleration factor 2). Functional images were acquired using an echoplanar imaging sequence with multi-echo (4) multiband (2; MEMB) acquisition. We used a multi-echo sequence in order to recover signal from the ATL, which is a region associated with semantic memory but suffers from susceptibility artefacts in gradient-echo fMRI[66–68]; we used multiband to compensate for the longer TR required for multi-echo acquisition. For the episodic task, volumes were acquired over three runs. For the semantic and the control tasks, volumes were acquired over two runs. Each volume contained 46 slices acquired in interleaved order within each excitation band, with a slice thickness of 3 mm and no interslice gap (TR of 1,792 ms; TE of 13, 25.85, 38.7 and 51.55 ms; flip angle of 75°; FOV of 192 mm × 192 mm; voxel size of 3 mm × 3 mm × 3 mm). Field maps for echoplanar imaging distortion correction were also collected (TR of 541 ms; TE of 4.92 ms; flip angle of 60°; FOV of 192 mm × 192 mm).

Preprocessing

All raw DICOM data were converted to nifti format using dcm2niix. The T1 data were processed using FSL (v5.0.11)[69–71] and subjected to the 'fsl_anat' function. This tool provides a general processing pipeline for anatomical images and involves the following steps (in order): (1) reorient images to standard space ('fslreorient2std'), (2) automatically crop image ('robustfov'), (3) bias-field correction ('fast'), (4) registration to Montreal Neurological Institute (MNI) space ('flirt' then 'fnirt'), (5) brain extraction (using fnirt warps) and (6) tissue-type segmentation ('fast'). Images warped to MNI space were visually inspected for accuracy.

The functional MEMB data were preprocessed using a combination of tools in FSL, AFNI (v18.3.03)[72] and a python package to perform TE-dependent analysis[73–75]. Images were despiked (3dDespike) and the slice-time corrected (3dTshift, to the middle slice) and realigned (3dvolreg) with AFNI before being submitted to the 'tedana' toolbox (max iterations of 100, max restarts of 10). Tenada takes the time series from all the collected TEs, decomposes the resulting data into components that can be classified as BOLD or non-BOLD based on their TE-dependence and then combines the echoes for each component, weighted by the estimated T2* in each voxel, and projects the noise components from the data[74,76]. Because the benefits of the ME-ICA denoising procedure were not yet established for task data when our Stage 1 report was accepted (although they were established in a recent paper, specifically designed to address this issue[52]), we stated then that we will analyse the data using two methods—combined and combined + ICA-denoise—and formally compare the results. This analysis showed that the combined + ICA-denoise procedure was consistently advantageous compared to the combined procedures. The full details and results of this analysis are included in Supplementary Analysis 2.

The resulting denoised and combined images were then averaged, and the mean coregistered to T1 (flirt) and warped to MNI space using fnirt warps and flirt transform. For whole-brain activation analyses (that is, when extracting the data-driven clusters, see below), smoothing was applied using an 8 mm full width at half maximum Gaussian kernel.

Data analysis

Trial classification. Logo-cue trials in all three tasks were classified into two response types of interest based on the response to that trial: (1) success: trials in which the associated brand from the study phase (episodic)/the actual associated brand (semantic)/the letters combination (control) was correctly recalled; (2) failure: all trials with a 'pass' response. Logo-cue trials and name-cue trials were further classified into detailed (success trials for which many details were retrieved/extracted) and not detailed (success trials for which few or no details were retrieved/extracted). Each success trial was then assigned into one of three categories, according to the correspondence between logo-cue and name-cue trials: (1) match(+): an entry that was classified as retrieved with many details in both the logo-cue trial and the name-cue trial; (2) match(−): an entry that was classified as retrieved with few details in both trials; and (3) mismatch: an entry that was classified as retrieved with many details in the logo-cue trials but with few details in the corresponding name-cue trial or vice versa. Trials that were given incorrect responses (false alarm trials), and any trials in which no response was collected during the scanning session were included in the general linear model (GLM) model ('Univariate analysis' section) but disregarded from the analyses. As was mentioned above, to minimize semantic processing in the episodic task, trials corresponding to logos that were classified as 'can name' in the debriefing session, regardless of whether or not the correct response was provided, were excluded from the analyses at the first instance. Nevertheless, to increase statistical power, we ran the analyses that determined the stopping criteria for the experiment ('Selection and definition of locations of interest' section) both with and without these trials, and because numerical trends remained the same, these trials were included in our main analyses.

General approach. Statistical analyses were performed in R and RStudio. Both behavioural and fMRI data were analysed using Bayesian mixed-models analyses that accommodate both within- and between-participant variability. This approach is particularly recommended for unbalanced data (an unequal number of trials in each condition[77]), which we have here due to the post hoc division of trials into successful and failure trials. Categorical predictors were dummy-coded. For binary outcomes (behavioural data), we used a logistic regression model with binomial family link function, whereas for continuous outcomes (fMRI data) we used a linear regression model with Gaussian family link function. We fitted all models (unless otherwise stated, in some of our exploratory analyses) across 36,000 samples of which the first 3,000 were warm-up samples, using the Bayesian regression models (BRMS) R package[78,79].

Behavioural analyses. After classifying the trials according to their responses, we analysed the behavioural data from the test phase of the episodic and semantic tasks as well as the data from the control task. Because the study phase is not comparable across tasks, these data were not subjected to statistical comparisons but are available on https://openneuro.org/datasets/ds004495 for completion and potential future use.

To analyse behavioural data from the test phase, trial outcomes (1 for success, 0 for failure) were submitted to a Bayesian logistic regression model, which included task (episodic, semantic, control) as a fixed factor and a participant-specific intercept and slope for this factor as the random part of the model. We fitted the model with the following formula, where '(… |x)' indicates random effects:

$$\text{Outcome} \approx \text{task} + (1 + \text{task}|\text{participant}).$$

Selection and definition of locations of interest. We used two definitions to identify locations of interest. The first was an a priori definition of networks, based on functional data from previous studies. For this definition, we identified two networks of interest: a core-recollection network that is applicable in various episodic tasks, regardless the nature of the recollected content[11,12], and a semantic network, comprising regions that show consistent activation in tasks that require semantic processing[80], combined with the ATL from a recent fMRI study using a paradigm with enhanced ATL signal coverage[81]. To maximize our ability to detect potential differences, we excluded any overlapping voxels that are shared between the two networks. These networks are shown in Fig. 2a. Note that each network includes several regions of non-contiguous voxels. This definition was used for setting the stopping criteria for data collection.

The second definition was based on data from the current study (that is, data-defined clusters). For this definition, we ran a whole-brain analysis and identified locations where the recall success effect (success-failure) in the two critical tasks (episodic and semantic) was greater than the success effect in the control task ($P < 0.05$ family-wise error cluster-level corrected, with voxel-level threshold at $P < 0.001$ uncorrected). Note that this contrast is orthogonal to the difference between episodic and semantic success effects, so it does not bias the subsequent comparisons of these two. The sample used for this analysis consisted of 32 participants, following the exclusion of 4 additional participants: 2 for which behavioural data for the control task were not available due to problems with the recordings and 2 for which MRI data for the control task were not recorded due to technical issues. Two clusters were obtained with this definition, shown in Fig. 2b.

Univariate analysis. The univariate analysis was conducted for logo-cue trials classified as 'success' and 'failure' trials ('Trial classification' section). It was conducted separately for a priori-defined networks and data-defined clusters.

SPM12 (www.fil.ion.ucl.ac.uk/spm) in Matlab (The MathWorks) was used to construct GLMs for each participant separately for each run and for each task. These first-level GLMs included two separate regressors for each response type of interest for logo-cue trials (success, failure), a regressor for excluded responses (filler trials, false alarms and unnameable targets; 'Trial classification' section) and a regressor for name-cue trials which are of no interest for this analysis. Predicted responses in these regressors were locked to the onset of stimulus presentation. The GLM further included a regressor for the verbal response (at the onset of the 'pair-associate?' slide) and a regressor for the motor response to the 'Many/Few?' slide (locked to the motor response, for both logo- and name-cue trials). Each of these regressors were generated with a delta function convolved with a canonical haemodynamic response function. Six participant-specific movement parameters were also included to capture residual movement-related artefacts. The GLM was fitted to the data in each voxel. The autocorrelation of the error was estimated using an AR(1)-plus-white-noise model, together with a set of cosines to high-pass the model and data to 1/128 Hz to remove low-frequency noise, fitted using restricted maximum likelihood. The estimated error autocorrelation was then used to 'prewhiten' the model and data, after which ordinary least squares were used to estimate the model parameters.

Group level analyses were conducted using mixed-effect models. Importantly, rather than averaged estimates across participant/condition, the mixed-models require estimation of the BOLD response for each trial. To get these estimations, we used the locations described above (that is, a priori-defined networks and data-defined clusters). For each participant, we extracted time series data from these locations using the first eigenvector across all voxels in each location and adjusting for effects of no interest such as those captured by the motion regressors (see first-level GLMs above). To estimate the BOLD response to each trial from this time series, we used the least-squares separate (LSS-N) approach[82,83], where 'N' is the number of conditions. LSS-N fits a separate GLM for each trial, with one regressor for the

trial of interest and one regressor for all other trials of each of the N conditions. Only single trials of the conditions of interest were estimated this way. LSS implements a form of temporal smoothness regularization on the parameter estimation[83]. The regressors were created by convolving a delta function at the onset of each stimulus with the haemodynamic response function. The parameters for the regressor of interest were then estimated using ordinary least squares, and the whole process was repeated for each separate trial of interest and for each location.

The resulting betas were submitted to a Bayesian linear mixed-model which included two locations (either predefined or data driven, depending on the analysis), the two critical tasks (episodic, semantic), response type (success, failure) and their interactions as fixed effects. The model further included participant-specific slopes for each factor and a participant-specific intercept (see discussion on estimation and convergence problems of the maximal random effects model[84,85]) as the random part of the model. The 'ROI' factor always included two locations. We fitted the model with the following formula, with a unit normal prior set for the intercept and for the mean effect of x:

$$\text{Betas} \approx \text{ROI} \times \text{task} \times \text{response type}$$
$$+(1 + \text{ROI} + \text{task} + \text{response type}|\text{participant}).$$

We then tested our prediction of greater recall success effect (success > failure) in the episodic task in the episodic network but greater recall success effect in the semantic task in the semantic network for a priori-defined networks. We also tested a two-tailed version of the same interaction between ROI and task for the data-defined clusters. The present of any such interaction involving two locations would support our hypothesis that the episodic and semantic systems are dissociable.

Similarity analysis. The similarity analysis was conducted for success trials in the two critical tasks (episodic, semantic) and was restricted to pairs of logo- and name-cue trials that corresponded to the same level of details (that is, classified as 'match(+)' or 'match(−)' trials; 'Trial classification' section). In addition, because pairs of logo-brand trials that correspond to the same event ('same trials') were always from the same block (run), each logo-cue trial was only correlated with an unassociated name-cue trial from the same block, to avoid confounding same/different with temporal lag between logo-cue and name-cue trials.

First-level GLMs included the same regressors as in the univariate analysis, with additional regressors for logo-cue and name-cue trials modelled as a 5 s boxcar function at the onset of the delay period. Similar to the univariate analysis, the similarity analysis was performed separately for each type of location definition. Nevertheless, instead of extracting a single time-course from each location, we extracted a time-course from each voxel within the locations. Moreover, the time of interest was the delay period where no perceptual information was presented (5 s following stimulus presentation, modelled as an epoch). For this analysis, we focused on the delay period rather than stimulus onset, as was done for the univariate analysis. Epoch data were extracted using the LSS-N approach described above. Following this extraction of single-trial time courses, successfully recalled logo-cue trials and their corresponding name-cue trials, were considered to be trials-of-interest if labelled as 'match' trials ('Trial classification' section). Other trials (mismatch, failure, false alarms or unnamable targets) were excluded.

Logo-brand similarity was then computed using Fisher -transformed correlation coefficients for (1) 'same' pairs: success logo-cue trials classified as detailed/non-detailed and the name-cue trial, which corresponds to the associated brand (that is, from the study phase in the episodic task or the actual associated brand in the semantic task), and (2) 'different' pairs: success logo-cue trials classified as detailed/non-detailed and a different (unassociated) name-cue trial corresponding to the same level of detailedness, and presented at the smallest time-lag, either before or after, from the associated brand-test trial. These similarity coefficients were submitted to a linear mixed-model with ROI, task (episodic, semantic), trial type (same, different) and their interactions, as the fixed part of the model, and with participant-specific slopes (for each factor) and intercept as the random factors with the same priors described above:

$$\text{Coefficient} \approx \text{ROI} \times \text{task} \times \text{trial type}$$
$$+(1 + \text{ROI} + \text{task} + \text{trial type}|\text{participant}).$$

We tested our prediction of greater similarity effect (same -different) in the episodic task in the episodic network but greater similarity effect in the semantic task in the semantic network (for a priori-defined networks). We also tested a two-tailed version of the same interaction between ROI and task for the data-defined ROIs. We rationalized that such pattern would indicate that semantic and episodic content is represented in dissociable brain regions.

Deviations from registered protocol

- For each participant, we intended that the two experimental sessions will take place 2–21 days apart. However, due to unpredictable circumstances (illness, unexpected leave, MRI scanner problems), the scans were further apart for four participants (27, 42, 47 and 63 days). For the rest, the time between sessions ranged between 2 and 18 days (mean of 7.72 days, s.d. of 4.83).
- Our Stage 1 report stated that for the control task, a block of ten practice trials will be provided before the scan. In practice, this task was straightforward and did not require much practice, and so a block of five practice trials was given instead.
- For the analysis of behavioural data, we registered the following model: outcome ≈ task × response type + (1 + task + response type|participant). In retrospect, however, we realized that this cannot be done because the dependent variable 'outcome' is identical to the 'response type' factor (that is, both indicate success/failure). We therefore removed the 'response type' factor from the model and used the following formula instead: outcome ≈ task + (1 + task|participant).
- In Stage 1, we referred to the episodic and semantic networks of interests that were defined based on previous literature as 'anatomical ROIs' and to regions that were selected based on current data as 'functional ROIs'. However, the episodic and semantic networks derived from previous literature were also defined functionally rather than anatomically. We therefore changed the terminology throughout the manuscript to 'a priori-defined networks' to reflect regions that were selected based on previous literature and 'data-defined clusters' or 'data-defined ROIs' to reflect clusters/regions selected based on current data. Note that some of the shared code uses the old terminology.
- In our Stage 1 report, we stated that to avoid an inflated false positive rate due to multiple comparisons, we will adjust the model's prior based on the number of comparisons that we perform. In practice, this was not required because our registered analysis only yielded two clusters, resulting in a single comparison.

Protocol registration

The approved Stage 1 protocol is available at https://osf.io/dm47y/.

Reporting summary

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

Data availability

Raw neuroimaging data, behavioural data and lab logs are available at https://openneuro.org/datasets/ds004495/. Processed data, pilot data and materials are available at https://osf.io/dm47y/.

Code availability

All code used for the study is available at https://osf.io/dm47y/.

References

1. Tulving, E. in *Organization of Memory* (eds Tulving, E. & Donaldson, W.) 381–403 (Academic Press, 1972).
2. Tulving, E. Memory and consciousness. *Can. Psychol. Can.* **26**, 1–12 (1985).
3. Nebes, R. D., Martin, D. C. & Horn, L. C. Sparing of semantic memory in Alzheimer's disease. *J. Abnorm. Psychol.* **93**, 321–330 (1984).
4. Hodges, J. R. & Patterson, K. Is semantic memory consistently impaired early in the course of Alzheimer's disease? Neuroanatomical and diagnostic implications. *Neuropsychologia* **33**, 441–459 (1995).
5. Vargha-Khadem, F. et al. Differential effects of early hippocampal pathology on episodic and semantic memory. *Science* **277**, 376–380 (1997).
6. Graham, K. S., Kropelnicki, A., Goldman, W. P. & Hodges, J. R. Two further investigations of autobiographical memory in semantic dementia. *Cortex* **39**, 729–750 (2003).
7. McKinnon, M. C., Black, S. E., Miller, B., Moscovitch, M. & Levine, B. Autobiographical memory in semantic dementia: Implications for theories of limbic-neocortical interaction in remote memory. *Neuropsychologia* **44**, 2421–2429 (2006).
8. Snowden, J., Griffiths, H. & Neary, D. Semantic dementia: autobiographical contribution to preservation of meaning. *Cogn. Neuropsychol.* **11**, 265–288 (1994).
9. Westmacott, R. & Moscovitch, M. The contribution of autobiographical significance to semantic memory. *Mem. Cognit.* **31**, 761–774 (2003).
10. Renoult, L., Irish, M., Moscovitch, M. & Rugg, M. D. From knowing to remembering: the semantic–episodic distinction. *Trends Cogn. Sci.* **23**, 1041–1057 (2019).
11. King, D. R., Chastelaine, M. de, Elward, R. L., Wang, T. H. & Rugg, M. D. Recollection-related increases in functional connectivity predict individual differences in memory accuracy. *J. Neurosci.* **35**, 1763–1772 (2015).
12. Thakral, P. P., Wang, T. H. & Rugg, M. D. Decoding the content of recollection within the core recollection network and beyond. *Cortex* **91**, 101–113 (2017).
13. Binder, J. R. & Desai, R. H. The neurobiology of semantic memory. *Trends Cogn. Sci.* **15**, 527–536 (2011).
14. Irish, M. & Vatansever, D. Rethinking the episodic–semantic distinction from a gradient perspective. *Curr. Opin. Behav. Sci.* **32**, 43–49 (2020).
15. Irish, M. in *The Cambridge Handbook of Imagination* (ed. Abraham, A.) 447–465 (Cambridge Univ. Press, 2020).
16. Burianova, H. & Grady, C. L. Common and unique neural activations in autobiographical, episodic, and semantic retrieval. *J. Cogn. Neurosci.* **19**, 1520–1534 (2007).
17. Burianova, H., McIntosh, A. R. & Grady, C. L. A common functional brain network for autobiographical, episodic, and semantic memory retrieval. *NeuroImage* **49**, 865–874 (2010).
18. Rajah, M. N. & McIntosh, A. R. Overlap in the functional neural systems involved in semantic and episodic memory retrieval. *J. Cogn. Neurosci.* **17**, 470–482 (2005).
19. Ryan, L., Cox, C., Hayes, S. M. & Nadel, L. Hippocampal activation during episodic and semantic memory retrieval: comparing category production and category cued recall. *Neuropsychologia* **46**, 2109–2121 (2008).
20. Vatansever, D., Smallwood, J. & Jefferies, E. Varying demands for cognitive control reveals shared neural processes supporting semantic and episodic memory retrieval. *Nat. Commun.* **12**, 2134 (2021).
21. Humphreys, G. F., Jung, J. & Lambon Ralph, M. A. The convergence and divergence of episodic and semantic functions across lateral parietal cortex. *Cereb. Cortex* **32**, 5664–5681 (2022).
22. Nyberg, L. et al. Common prefrontal activations during working memory, episodic memory, and semantic memory. *Neuropsychologia* **41**, 371–377 (2003).
23. St-Laurent, M., Abdi, H., Burianová, H. & Grady, C. L. Influence of aging on the neural correlates of autobiographical, episodic, and semantic memory retrieval. *J. Cogn. Neurosci.* **23**, 4150–4163 (2011).
24. Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U. & Bäckman, L. Memory aging and brain maintenance. *Trends Cogn. Sci.* **16**, 292–305 (2012).
25. Patterson, K. & Lambon Ralph, M. A. in *Neurobiology of Language* (eds. Hickok, G. & Small, S. L.) 765–775 (Academic Press, 2016); https://doi.org/10.1016/B978-0-12-407794-2.00061-4
26. Bonnici, H. M., Richter, F. R., Yazar, Y. & Simons, J. S. Multimodal feature integration in the angular gyrus during episodic and semantic retrieval. *J. Neurosci.* **36**, 5462–5471 (2016).
27. Rugg, M. D. & Wilding, E. L. Retrieval processing and episodic memory. *Trends Cogn. Sci.* **4**, 108–115 (2000).
28. Wetzels, R., Grasman, R. P. P. P. & Wagenmakers, E.-J. An encompassing prior generalization of the Savage–Dickey density ratio. *Comput. Stat. Data Anal.* **54**, 2094–2102 (2010).
29. Frisby, S. L., Halai, A. D., Cox, C. R., Lambon Ralph, M. A. & Rogers, T. T. Decoding semantic representations in mind and brain. *Trends Cogn. Sci.* **27**, 258–281 (2023).
30. Cabeza, R. et al. Attention-related activity during episodic memory retrieval: a cross-function fMRI study. *Neuropsychologia* **41**, 390–399 (2003).
31. Shallice, T. et al. Brain regions associated with acquisition and retrieval of verbal episodic memory. *Nature* **368**, 633–635 (1994).
32. Rugg, M. D., Fletcher, P. C., Frith, C. D., Frackowiak, R. S. J. & Dolan, R. J. Differential activation of the prefrontal cortex in successful and unsuccessful memory retrieval. *Brain* **119**, 2073–2083 (1996).
33. Henson, R. N., Shallice, T. & Dolan, R. J. Right prefrontal cortex and episodic memory retrieval: a functional MRI test of the monitoring hypothesis. *Brain J. Neurol.* **122**, 1367–1381 (1999).
34. Zahn, R. et al. Recovery of semantic word processing in global aphasia: a functional MRI study. *Cogn. Brain Res.* **18**, 322–336 (2004).
35. Raposo, A., Moss, H. E., Stamatakis, E. A. & Tyler, L. K. Repetition suppression and semantic enhancement: An investigation of the neural correlates of priming. *Neuropsychologia* **44**, 2284–2295 (2006).
36. Brambati, S. M., Benoit, S., Monetta, L., Belleville, S. & Joubert, S. The role of the left anterior temporal lobe in the semantic processing of famous faces. *NeuroImage* **53**, 674–681 (2010).
37. Noonan, K. A., Jefferies, E., Visser, M. & Lambon Ralph, M. A. Going beyond inferior prefrontal involvement in semantic control: evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. *J. Cogn. Neurosci.* **25**, 1824–1850 (2013).
38. D'Mello, A. M., Centanni, T. M., Gabrieli, J. D. E. & Christodoulou, J. A. Cerebellar contributions to rapid semantic processing in reading. *Brain Lang.* **208**, 104828 (2020).
39. D'Mello, A. M., Turkeltaub, P. E. & Stoodley, C. J. Cerebellar tDCS modulates neural circuits during semantic prediction: a combined tDCS-fMRI study. *J. Neurosci.* **37**, 1604–1613 (2017).

40. Stoodley, C. J. The cerebellum and cognition: evidence from functional imaging studies. *Cerebellum* **11**, 352–365 (2012).
41. Gatti, D., Vecchi, T. & Mazzoni, G. Cerebellum and semantic memory: a TMS study using the DRM paradigm. *Cortex* **135**, 78–91 (2021).
42. Henson, R. Forward inference using functional neuroimaging: dissociations versus associations. *Trends Cogn. Sci.* **10**, 64–69 (2006).
43. Gould, R. L. et al. Brain mechanisms of successful compensation during learning in Alzheimer disease. *Neurology* **67**, 1011–1017 (2006).
44. Geerligs, L., Renken, R. J., Saliasi, E., Maurits, N. M. & Lorist, M. M. A brain-wide study of age-related changes in functional connectivity. *Cereb. Cortex* **25**, 1987–1999 (2015).
45. Geerligs, L., Maurits, N. M., Renken, R. J. & Lorist, M. M. Reduced specificity of functional connectivity in the aging brain during task performance. *Hum. Brain Mapp.* **35**, 319–330 (2014).
46. Rubin, D. C. A conceptual space for episodic and semantic memory. *Mem. Cognit.* **50**, 464–477 (2022).
47. Addis, D. R. & Szpunar, K. K. Beyond the episodic–semantic continuum: the multidimensional model of mental representations. *Philos. Trans. R. Soc. B* **379**, 20230408 (2024).
48. Tanguay, A. F. et al. The shared and unique neural correlates of personal semantic, general semantic, and episodic memory. *eLife* **12**, e83645 (2023).
49. Curot, J. et al. Complex memories induced by intracranial electrical brain stimulation are related to complex networks. *Cortex* **183**, 349–372 (2025).
50. Maguire, E. A., Kumaran, D., Hassabis, D. & Kopelman, M. D. Autobiographical memory in semantic dementia: A longitudinal fMRI study. *Neuropsychologia* **48**, 123–136 (2010).
51. Hodges, J. R. & Patterson, K. Semantic dementia: a unique clinicopathological syndrome. *Lancet Neurol.* **6**, 1004–1014 (2007).
52. Halai, A. D., Henson, R. N., Finoia, P. & Correia, M. M. Comparing the effect of multi-gradient echo and multi-band fMRI during a semantic task. *Imaging Neurosci.* **3**, IMAG.a.1043 (2025).
53. Gilmore, A. W., Agron, A. M., González-Araya, E. I., Gotts, S. J. & Martin, A. A comparison of single- and multi-echo processing of functional MRI data during overt autobiographical recall. *Front. Neurosci.* **16**, 854387 (2022).
54. Allen, T. A. & Fortin, N. J. The evolution of episodic memory. *Proc. Natl Acad. Sci. USA* **110**, 10379–10386 (2013).
55. Tulving, E. & Markowitsch, H. J. Episodic and declarative memory: role of the hippocampus. *Hippocampus* **8**, 198–204 (1998).
56. Eichenbaum, H. & Fortin, N. J. Bridging the gap between brain and behavior: cognitive and neural mechanisms of episodic memory. *J. Exp. Anal. Behav.* **84**, 619–629 (2005).
57. Moscovitch, M., Cabeza, R., Winocur, G. & Nadel, L. Episodic memory and beyond: the hippocampus and neocortex in transformation. *Annu. Rev. Psychol.* **67**, 105–134 (2016).
58. Hassabis, D., Kumaran, D. & Maguire, E. A. Using imagination to understand the neural basis of episodic memory. *J. Neurosci.* **27**, 14365–14374 (2007).
59. Reas, E. T. & Brewer, J. B. Retrieval search and strength evoke dissociable brain activity during episodic memory recall. *J. Cogn. Neurosci.* **25**, 219–233 (2013).
60. Kim, H., Daselaar, S. M. & Cabeza, R. Overlapping brain activity between episodic memory encoding and retrieval: roles of the task-positive and task-negative networks. *NeuroImage* **49**, 1045–1054 (2010).
61. Tibon, R., Fuhrmann, D., Levy, D. A., Simons, J. S. & Henson, R. N. Multimodal integration and vividness in the angular gyrus during episodic encoding and retrieval. *J. Neurosci.* **39**, 4365–4374 (2019).
62. Schönbrodt, F. D. & Wagenmakers, E.-J. Bayes factor design analysis: planning for compelling evidence. *Psychon. Bull. Rev.* **25**, 128–142 (2018).
63. Keizer, K. Impulse noise addition. *MATLAB* https://uk.mathworks.com/matlabcentral/fileexchange/22141-impulse-noise-addition (2025).
64. Drucker, D. M., Kerr, W. T. & Aguirre, G. K. Distinguishing conjoint and independent neural tuning for stimulus features with fMRI adaptation. *J. Neurophysiol.* **101**, 3310–3324 (2009).
65. Persichetti, A. S., Aguirre, G. K. & Thompson-Schill, S. L. Value is in the eye of the beholder: early visual cortex codes monetary value of objects during a diverted attention task. *J. Cogn. Neurosci.* **27**, 893–901 (2014).
66. Halai, A. D., Welbourne, S. R., Embleton, K. & Parkes, L. M. A comparison of dual gradient-echo and spin-echo fMRI of the inferior temporal lobe. *Hum. Brain Mapp.* **35**, 4118–4128 (2014).
67. Halai, A. D., Parkes, L. M. & Welbourne, S. R. Dual-echo fMRI can detect activations in inferior temporal lobe during intelligible speech comprehension. *NeuroImage* **122**, 214–221 (2015).
68. Caballero-Gaudes, C. & Reynolds, R. C. Methods for cleaning the BOLD fMRI signal. *NeuroImage* **154**, 128–149 (2017).
69. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL. *NeuroImage* **62**, 782–790 (2012).
70. Smith, S. M. et al. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* **23**, S208–S219 (2004).
71. Woolrich, M. W. et al. Bayesian analysis of neuroimaging data in FSL. *NeuroImage* **45**, S173–S186 (2009).
72. Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–173 (1996).
73. DuPre, E. et al. ME-ICA/Tedana: 0.0.9a. *Zenodo* https://doi.org/10.5281/zenodo.3786890 (2020).
74. Kundu, P., Inati, S. J., Evans, J. W., Luh, W.-M. & Bandettini, P. A. Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *NeuroImage* **60**, 1759–1770 (2012).
75. Kundu, P. et al. Integrated strategy for improving functional connectivity mapping using multiecho fMRI. *Proc. Natl Acad. Sci. USA* **110**, 16187–16192 (2013).
76. Poser, B. A., Versluis, M. J., Hoogduin, J. M. & Norris, D. G. BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: parallel-acquired inhomogeneity-desensitized fMRI. *Magn. Reson. Med.* **55**, 1227–1235 (2006).
77. Tibon, R. & Levy, D. A. Striking a balance: analyzing unbalanced event-related potential data. *Front. Psychol.* **6**, 555 (2015).
78. Bürkner, P.-C. brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28 (2017).
79. Bürkner, P.-C. Advanced Bayesian multilevel modeling with the R package brms. *R J.* **10**, 395–411 (2018).
80. Binder, J. R., Desai, R. H., Graves, W. W. & Conant, L. L. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* **19**, 2767–2796 (2009).
81. Humphreys, G. F., Hoffman, P., Visser, M., Binney, R. J. & Ralph, M. A. L. Establishing task- and modality-dependent dissociations between the semantic and default mode networks. *Proc. Natl Acad. Sci. USA* **112**, 7857–7862 (2015).
82. Mumford, J. A., Davis, T. & Poldrack, R. A. The impact of study design on pattern estimation for single-trial multivariate pattern analysis. *NeuroImage* **103**, 130–138 (2014).
83. Abdulrahman, H. & Henson, R. N. Effect of trial-to-trial variability on optimal event-related fMRI design: implications for beta-series correlation and multi-voxel pattern analysis. *NeuroImage* **125**, 756–766 (2016).

84. Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* **68**, 255–278 (2013).
85. Bates, D., Kliegl, R., Vasishth, S. & Baayen, H. Parsimonious mixed models. Preprint at https://arxiv.org/abs/1506.04967 (2018).

Acknowledgements

R.T. was supported by a British Academy Postdoctoral Fellowship (grant no. SUAI/028 RG94188). G.H. was supported by an MRC Programme grant to Matt Lambon Ralph (grant no. MR/R023883/1) and MRC intramural funding (grant no. MC_UU_00005/18). J.A.Q. was supported by the Gates Cambridge Trust PhD studentship. R.H. and A.G. were supported by a UK Medical Research Council grant (grant no. SUAG/010 RG91365). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We are grateful to A. Halai for advice on acquisition and preprocessing of multi-echo data, to L. Johnson for help with data collection and to D. Suchowski and I. Leitner for help with figure preparation.

Author contributions

R.T. directed the project together with R.H., designed and planned the project, collected the data, carried out the analyses and wrote the paper with input from all authors. A.G. consulted on design and analyses, helped with data collection and provided input on the analyses and the paper. G.H. consulted on design and analyses and provided input on the analyses and the paper. J.A.Q. designed and executed sample size determination and provided input on the analyses and the paper. R.H. directed the project together with R.T., consulted on design and analyses and provided input on the analyses and the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-025-02390-4.

Correspondence and requests for materials should be addressed to Roni Tibon.

Peer review information *Nature Human Behaviour* thanks Louis Renoult, Deniz Vatansever and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.