Selective retrieval revealed by electroencephalographic (EEG) neural pattern reinstatement

Moccia, A.[1], Plummer, M., Morcom, A. M[1].

[1]University of Sussex

Keywords: Episodic memory, Recollection, EEG, Neural Pattern Reinstatement, Pre-retrieval Control

Acknowledgements
This work was supported by a BIAL Foundation grant to A.M. Morcom (ref.: 169-18).

Conflict of interest
The authors have no financial or non-financial interests to disclose.

Data availability statement
Data and analysis scripts are uploaded to the Open Science Foundation repository and will be made publicly available upon publication in a scientific journal. Raw data, task materials and scripts are publicly available and can be accessed here: https://osf.io/erjdu/ and https://osf.io/dnz9e/ (for Experiment 1 and Experiment 2, respectively).

Author contributions
A. Moccia and A.M. Morcom developed the study concept and design. Testing and data collection were performed by A. Moccia. M. Plummer, A. Moccia and A. M. Morcom designed the analysis strategy. A. Moccia, A. M. Morcom and M. Plummer performed data analyses. A. Moccia drafted the manuscript with M. Plummer's input to the Methods. A.M. Morcom and A. Moccia provided critical revisions. All authors approved the final version of the manuscript for submission.

Ethics approval.
All procedures performed in these studies were in line with the principles of the Declaration of Helsinki. Approval was granted by the Psychology Research Ethics Committee at the University of Edinburgh, ref.: 135-1819/1 (Experiment 1) and 300-1819/1 (Experiment 2).

Abstract

For past experiences to guide our actions we need to retrieve the relevant memories. Here, we used electroencephalography (EEG) to investigate how a memory is selected for retrieval from a large store of mnemonic traces, and to evaluate how selection operates during the retrieval cascade. We analyzed data from two studies in which people studied objects in picture or auditory word formats, and later recalled them using either written words (Experiment 1, N=28) or line drawings (Experiment 2, N=28) as retrieval cues. We used multivariate decoding to quantify the reinstatement of study phase neural patterns when people successfully identified items that had been studied in the format currently designated as targets, versus non-targeted items. Neural reinstatement emerged by 500 ms post-stimulus, as did the established left parietal event-related potential (ERP) signature of recollection. Reinstatement was also target-selective (greater for targets than non-targets) when test cues overlapped more with targets, as had previously been shown for the left parietal ERP (Moccia and Morcom, 2021). In contrast, when cues overlapped more with non-targets, neural reinstatement was non-selective or reversed, unlike the left parietal ERP. We also tested for goal-directed reinstatement prior to retrieval cues, hypothesized to drive selection. When words were cues there was strong evidence of this proactive reinstatement, but it was not detected when pictures were cues. Together, the data suggest that selection can act at multiple stages of recovery of a memory and depends on both external cues and goal-directed control.

Keywords: Episodic memory, Recollection, EEG, Neural Pattern Reinstatement, Pre-retrieval Control

Selective retrieval revealed by electroencephalographic (EEG) neural pattern reinstatement

Introduction

Our brains store memories of a vast number of past experiences. To use these experiences to guide behavior, we must somehow select the relevant memories for retrieval. Electrophysiological studies in humans have provided compelling evidence that selection can act before memories are retrieved (Herron & Rugg, 2003; Mecklinger, 2010; Rosburg & Mecklinger, 2017). However, little is known about when selection operates in the retrieval processing cascade, or how it is achieved. Here, we used time-resolved multivariate decoding of scalp-recorded electroencephalographic (EEG) patterns to investigate the neural dynamics of selective recollection and the goal representations proposed to enable it.

Current evidence for pre-retrieval selection comes mainly from event-related potential (ERP) studies measuring the left parietal 'old/new' effect, an established index of successful episodic retrieval. This ERP occurs about 500-800 ms after a retrieval test cue and is typically larger when items are recollected rather than only familiar (see Friedman & Johnson, 2000; Rugg & Curran, 2007). It is also larger when more detail is recollected (Vilberg & Rugg, 2007), and when recollection is more precise (Murray et al., 2015). Selective retrieval is shown when the left parietal ERP is larger for currently targeted than non-targeted items (Herron & Rugg, 2003; Rosburg & Mecklinger, 2017). A strength of this electrophysiological approach is its high temporal precision. We know that the left parietal ERP onsets at around the same time as intracranial recollection signals (Staresina & Wimber, 2019), so any modulation of this ERP can be inferred to be due to processing taking place before or during retrieval. However, the left parietal ERP is not thought to index the initial stages of recollection, but likely reflects activity in lateral and/or medial parietal cortex supporting integrative processing downstream in the retrieval cascade (Bergström et al., 2013; Rugg & Vilberg, 2013).

Recollection is thought to occur when a cue partially reinstates neural patterns associated with a stored memory trace (Alvarez & Squire, 1994; Norman & O'Reilly, 2003; Treves & Rolls, 1994). This partial cue then triggers pattern completion by the hippocampus, leading to the reinstatement of distributed neural patterns that represented the original experience (Buckner & Wheeler, 2001; Damasio, 1989; Kosslyn, 1994). Human functional magnetic resonance imaging (fMRI) studies have provided strong support for a role of cortical reinstatement in successful recollection (Johnson & Rugg, 2007; Polyn, 2005; Staresina, Henson, et al., 2012; and Buckner & Wheeler, 2001; Danker & Anderson, 2010 for reviews). Preliminary evidence suggests that this reinstatement can occur selectively, meaning that it is stronger for 'targeted' events or event features of current interest (Favila et al., 2018; Kuhl et al., 2013; McDuff et al., 2009; Wimber et al., 2015). However, due to the very low time resolution of fMRI it is difficult to separate selective recollection from later goal-dependent processes operating on the products of recollection. Initial studies applying multivariate decoding approaches to electrophysiological data have begun to reveal the representational dynamics of recollection (Jafarpour et al., 2014; Johnson et al., 2015; Wimber et al., 2012), but have not yet clarified the point at which selection occurs.

Reinstated cortical representations have been hypothesized to support the vivid and multi-dimensional experience referred to as "reliving the past" (Danker & Anderson, 2010; Tulving, 1983; Wheeler et al., 2000). Alternatively, neural reinstatement may be necessary but not sufficient for recollection, since it can be just as strong when people fail to recollect as when they succeed (Thakral et al., 2015, 2017). Initial reinstatement may be an automatic consequence of pattern completion (Moscovitch, 2008) and subject to further, constructive processing (Damasio, 1989; Favila et al., 2018; Linde-Domingo et al., 2019). We assumed a simple model of the retrieval cascade in which selection may operate at two non-exclusive

stages. Bottom-up selection assumes that target memories are selected at the point of pattern completion, so non-target memories are not reinstated. If top-down processing contributes nothing further to selection, reinstatement will follow the same pattern of selectivity as the left parietal ERP. In contrast, top-down retrieval selection assumes that both target and non-target information is reinstated but that top-down, goal-dependent processing amplifies target information. If this top-down mechanism operates alone, reinstatement will show a non-selective pattern when the left parietal ERP is selective for target memories.

Control over memory selection at retrieval means not relying only on external cues to trigger recollection. Numerous studies have shown goal-related ERP effects during retrieval attempts for unstudied (new) items (see Mecklinger, 2010). Goal-related EEG activity before retrieval cues are presented can also predict recollection success (Herron & Evans, 2018; Kerrén et al., 2021). The longstanding encoding specificity principle (Tulving & Thomson, 1973) suggests a mechanism for this pre-retrieval control: internal reinstatement of contextual information stored in sought-for memory traces. This proposal has been supported by indirect behavioural evidence (Smith & Vela, 2001). More direct neural evidence for internal reinstatement comes from both fMRI and EEG studies that have shown reactivation of neural patterns from targeted studied events. Goal-related neural activity has been detected prior to cued retrieval attempts in several studies, but does not directly show proactive reinstatement of study neural patterns before retrieval unfolds (Bramão et al., 2022; Bramão & Johansson, 2018; Manning et al., 2011; McDuff et al., 2009). This evidence is complemented by findings that study phase neural patterns reappear in the first second or so of memory search in free recall tasks, although such patterns are difficult to distinguish from those accompanying initial recollection (Kragel et al., 2021; Polyn, 2005; Polyn et al., 2012; see also Bramão et al., 2022; Bramão & Johansson, 2018).

Here, we examined two existing datasets for which the pattern of left parietal ERP target selectivity was known. In the previous report, the left parietal ERP was larger for targets than non-targets when the retrieval cues had a high degree of overlap with the targeted material (auditory words in Experiment 1, and pictures in Experiment 2; Moccia & Morcom, 2021). To assess the locus of selection, we investigated whether neural reinstatement is target-selective in the same task conditions in which the left parietal ERP is selective. We also tested the prediction that internal reinstatement occurs prior to retrieval.

Materials and Methods

Participants

Data from 56 right-handed participants were analyzed, N = 28 in Experiment 1 (20 female, 8 male, age M = 22.79 years, SD = 4.14) and N = 28 in Experiment 2 (20 female, 8 male, age M = 24.57 years, SD = 3.71). Data from one further participant in Experiment 2 were excluded due to an insufficient number of artefact-free trials. All reported having normal or corrected to normal vision and hearing, being in good health, not taking medication that might affect cognition, and being very fluent in English. Sample sizes were determined *a priori* for comparisons of left parietal ERPs (see Moccia & Morcom, 2021). The Psychology Research Ethics Committee at the University of Edinburgh approved the experiments, ref.: 135-1819/1 and 300-1819/1. Participants were students who were compensated with university credits or cash.

Materials

The same 262 pictures and names of common objects were used as stimuli in both experiments (Fig. 1). Of these, 240 served as critical items and 12 were used as fillers at the beginning of each study and test list, and 12 in practice lists. At study, objects appeared as either colored pictures or as auditory words spoken by an English native male voice. At test, memory was probed with visual words in Experiment 1, or grey-scale line drawings in

Experiment 2 (for further detail see Moccia & Morcom, 2021). The critical items were divided into 6 sets of 40 objects. For each of the two study-test cycles, one set of pictures and one of auditory words were combined to create a study list of 80 items. The corresponding visual words (Experiment 1) or line drawings (Experiment 2) were then combined with a third set of new items to create the test list of 120 items. For each study-test cycle, half the items studied as pictures, half those studied as auditory words, and half the new items were allocated to the first test block, and the remainder to the second test block. In total, there were 80 critical targets, 80 critical non-targets, and 80 critical new items. Item presentation order was determined randomly within each study and test list.

Procedure

The EEG was recorded in two experiments where participants performed a recognition exclusion task in which retrieval goals are manipulated (Jacoby, 1991). Selective recollection is quantified by comparing neural signatures of successful retrieval elicited by target versus non-target items (Herron & Rugg, 2003). The task involved two study-test cycles (Figure 1).

Study phase. Participants saw pictures or heard auditory words and were instructed to learn the items for a subsequent memory test, while rating their pleasantness on a 4-point scale from “very pleasant” to “not pleasant”. On each trial, a preparatory cue (pre-cue) presented for 1,000 ms signaled the format of the upcoming item, either a yellow asterisk ‘*’ or a blue lowercase ‘o’ (allocation to pictures and auditory words was counterbalanced). This was followed by a blank screen for 100 ms, then the stimulus for 1,000 ms. Afterwards, a 1,500 ms red fixation cross preceded the word “RESPOND”, shown at the center for up to 3,000 ms and during which participants responded via keyboard presses. A 100 ms blank screen separated each trial.

Test phase. Each study phase was followed by a brief pen-and-paper distractor task, then two memory test blocks with different target designations: in each, studied objects previously presented in one format were designated as targets (Target-Audio or Target-Pictures). For example, in the Target-Audio block, participants were instructed to answer ‘yes’ to an item if they remembered hearing the object name in the preceding study phase and ‘no’ to all other items. Target designation was also signaled on each trial using the same preparatory symbols as at study. In Experiment 1, retrieval cues were visual words, and in Experiment 2, they were grey-scale line drawings, in both cases shown at the center of the screen. Timing of test trials were as follows: stimuli were preceded by pre-cues for 500 ms and a black fixation for 1,800 ms. Then, cues were shown for 3,000 ms followed by a red fixation for 500 ms in Experiment 1, and 500 ms followed by a 3,000 ms fixation in Experiment 2. Participants’ responses were recorded during stimulus presentation in Experiment 1 and during fixation in Experiment 2. The order of the Target-Picture and Target-Audio blocks and the mapping of the keypress responses to left and right hands were counterbalanced across participants.

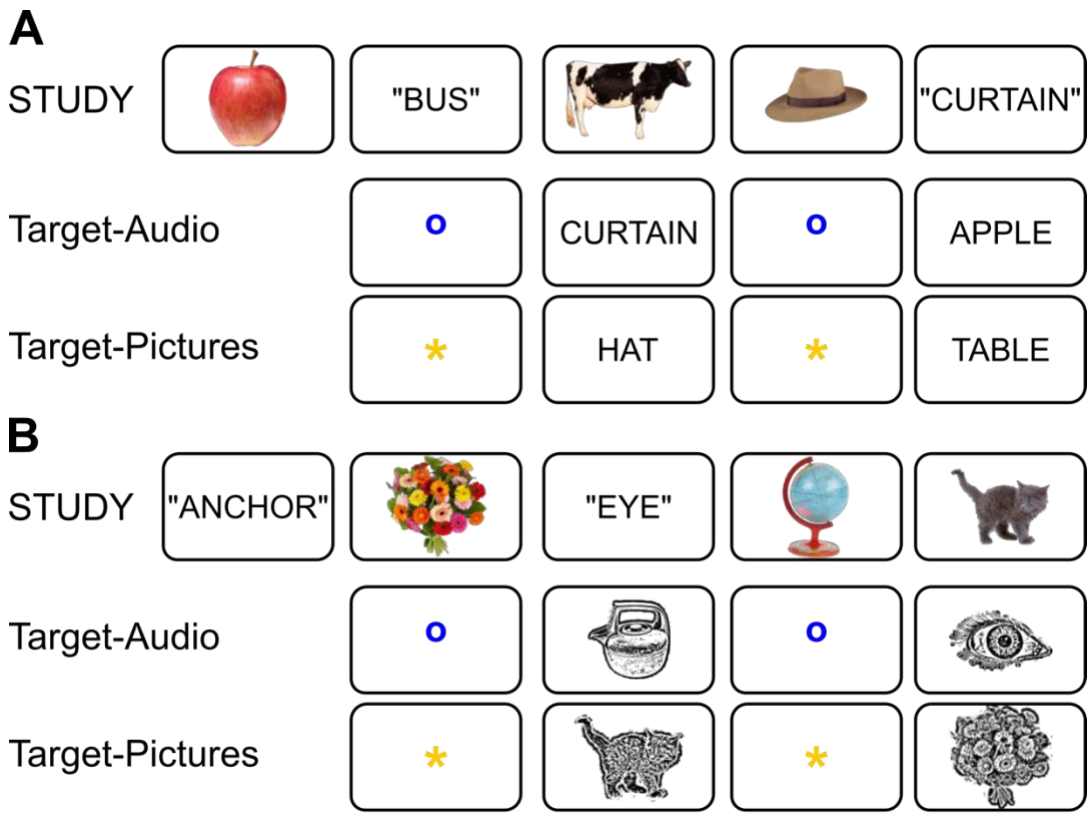

Figure 1. Experimental paradigm. A) Experiment 1 (visual word test cues); B) Experiment 2 (line drawing test cues). In both experiments, participants studied a single block of pictures intermixed with auditory words. In each of the test blocks, either items studied as pictures or auditory words were targets (see Materials and Procedure).

EEG recording and pre-processing

EEG data were recorded with a BioSemi Active Two AD-box with 24-bit signal digitization from 64 active silver/silver chloride electrodes in an elastic cap using the extended International 10-20 system configuration. Vertical and horizontal eye movements (electrooculogram; EOG) were recorded with bipolar electrodes above and below the right eye and on the outer canthi. EEG and EOG signals were acquired continuously at a 1024-Hz with amplifier bandwidth of 0±208 Hz (3 dB), referenced to a CMS electrode. EEG data were preprocessed using EEGLAB and MATLAB R2018a. A 0.1-40 Hz Hamming windowed-sinc FIR and a 50 Hz notch filter for line noise were applied after re-referencing the raw signal to linked mastoid electrodes. Following automatic rejection of gross artefacts, the data were partitioned into epochs including pre-cue and stimulus, time-locked to the stimulus onset (4,500 ms epochs for study and 6,700 ms for test). Vertical and horizontal EOG artefacts were corrected using Independent Component Analysis (ICA) with manual removal of components. Further details of recording and initial preprocessing are reported in Moccia & Morcom (2021). To increase signal-to-noise for multivariate decoding, we also smoothed the ERP data using a 20 ms full-width half-maximum Gaussian kernel.

Multivariate decoding analyses

To identify neural reinstatement, we used multivariate decoding analysis to assess the similarity between EEG neural patterns at study and at test, focusing on neural patterns that

differentiated memories of hearing auditory words and seeing pictures at study. We used linear discriminant analysis (LDA) with shrinkage regularization performed independently for each participant. MATLAB code (version 2019b) was custom-written using LDA functions adapted from Linde-Domingo et al. (2019).

Initial decoding of study phase data revealed significant decoding of items' study format throughout the stimulus presentation time-window (see next paragraph), which informed time-window selection for test phase decoding. We trained the LDA classifier to discriminate picture and audio trials using a leave-one-trial-out procedure (Linde-Domingo et al., 2019), training on all trials but one and testing on the left-out trial. Classifier features were the raw EEG amplitudes from the 64 scalp electrodes. Before decoding, data were downsampled to approximately 8 ms time bins. Multivariate noise normalization was applied to single trial data (Guggenmos et al., 2018; Linde-Domingo et al., 2019). Trials were subsampled to ensure equal numbers per class and the outcomes averaged over 12 iterations. Next, we assessed reinstatement of study phase neural patterns by cross classifying study and test data, pairing each training (study phase) with each test (test phase) time point to yield a matrix of values. The LDA classifier was trained to discriminate between pictures and auditory words at study, then tested on ERPs to test cues on trials attracting correct responses. This allowed us to decode information about study format from neural patterns during the memory test when the retrieval cues – visual words in Experiment 1 or line drawings in Experiment 2 – were not indicative of study format. Before decoding, study and test data were downsampled to approx. 8 ms time bins. ERPs were also baseline-corrected using a 200 ms pre-stimulus period.

To test whether distinct information from the study phase was reinstated for items studied as pictures versus auditory words and to compare target to non-target reinstatement, we measured neural patterns from four different *a priori* study phase time windows. The first two corresponded to time at which (early) perceptual and (later) conceptual reinstatement was previously observed during retrieval of pictures using word cues (Linde-Domingo et al., 2019). These were designated as the S1 (136-232 ms) and S2 (240-360 ms) study windows, rounded to fit our time bins. The second two windows were derived from the study phase decoding analysis of Experiment 1, which showed a peak corresponding closely to the two early windows followed by a plateau of above-zero decoding persisting until almost 1,000 ms after stimulus onset (Supplemental Figure S1). These were designated as the S3 (368-672 ms) and S4 (680-984 ms) study windows. The test phase data were also divided into an early and late time window (203-487 ms and 492-776 ms): the latter window corresponds to the typical timing of the left parietal ERP (Rugg & Curran, 2007).

To test for goal-related internal reinstatement, we then examined reinstatement of the same study phase neural patterns during the pre-cue phase of the test trials. An LDA classifier was again trained on study phase data, then tested on test phase ERPs time-locked to the pre-stimulus preparatory cues. The study phase time-window was the same as for the retrieval cue analysis (146-1,000 ms; differences reflecting subsampling to 8 versus 15 ms bins) and the test phase window was 190-2,300 ms after the preparatory cue. Study and test phase data for the pre-cue analysis were downsampled to approx. 15 ms time bins to accommodate the longer test data epoch during pre-cue presentation. To enable detection of state-related activity these data were not baseline-corrected.

In all LDA analyses, we extracted a decision (*d*) value for each trial and time-point that indicates the distance to the decision boundary dividing the two classes in a high-dimensional space. The sign of *d* indicates the category the data-point was classified as (+ for pictures and - for auditory words). Following Linde-Domingo et al. (2019), we inverted this sign for auditory word trials so correct classification was always associated with a positive *d* value. The resulting value then indexed classifier fidelity: how confidently the classifier

assigned that trial's neural pattern to its class. We summed the *d* values over trials in each class (condition), so fidelity values were always more positive the stronger the decoding of study phase neural patterns.

Statistical analysis

We examined mean *d* values for the four study windows and two test windows using analysis of variance (ANOVA) using the aov_4 function in the afex package (Singmann et al., 2022) in R (version 3.5.4, https://www.r-project.org). Degrees of freedom and *p*-values were Greenhouse-Geisser corrected where relevant and alpha set at .05, with Benjamini and Hochberg, (1995) corrections for multiple *post hoc* analyses. For ANOVAs, only significant effects are reported (but see Supplemental Tables S2-S7 for complete results). Two statistical outliers in Experiment 1 and two in Experiment 2 (± 3 SD from the mean) did not alter the pattern of statistical findings, so analyses of the full samples are reported but we used pairwise and one-sample Wilcoxon signed-rank tests for *post hoc* analysis. For exploratory analyses we used a cluster permutation multiple comparisons correction (Maris & Oostenveld, 2007) over study x test maps. After identifying significant clusters for each contrast with 1-tailed paired-sample *t* tests comparing fidelity values to 0 at alpha = .05, we computed cluster mass as the sum of *t*-values per cluster. Statistical significance (at 1-tailed alpha = .05) was then calculated by comparing the cluster mass against its permutation distribution, based on 1,000 LDA iterations with randomly shuffled training trial class labels.

Results

Behavioral results

In both Experiment 1, when test cues were words, and in Experiment 2, when test cues were line drawings, memory tended to be better for items studied as pictures, whether tested as targets or non-targets. Response times (RTs) were also faster for targets studied as pictures than audios in Experiment 1, and for all trial types in Experiment 2 (see supplementary Table S1; for statistical analyses see Moccia and Morcom, 2021).

Decoding reinstated neural patterns

To quantify neural reinstatement in each retrieval condition we tested the trained LDA classifiers on trial data for studied items (targets and/or non-targets) versus new items. On new trials, no information should be retrieved. Decoding was run within test blocks, comparing 1) targets studied as pictures versus new items (in the Target-Picture block); 2) targets studied as auditory words versus new items (in the Target-Audio block); 3) non-targets studied as auditory words versus new items (in the Target-Picture block); 4) Non-targets studied as pictures versus new items (in the Target-Audio block). In each case the studied items were coded according to their study format and the new items were coded as the other study format: for example, for contrast 1), we tested the classifier with picture targets coded as pictures and new items as the audio class, and for contrast 3) picture block non-targets were coded as audio items and new items as the picture class. The results of the analyses for each studied item condition versus new are reported in the Supplementary Material (Supplementary LDA results: Test phase decoding). Our main analyses focused on two sets of questions, concerning format-specificity and target-selectivity of reinstatement during successful retrieval.

Reinstated neural patterns are format-specific

We first assessed whether the information carried in reinstated neural patterns at retrieval differed according to whether the object was studied in a picture or audio format, as would be expected if reinstatement involves format-specific cortical networks (Danker &

Anderson, 2010). We therefore asked whether reinstated information about auditory and pictorial study episodes derived from different temporal stages of study phase processing. Since our first two study phase time windows were previously associated with perceptual (S1) versus semantic (S2) processing of picture-word pairs, we predicted that study neural patterns distinguishing picture and audio format items would derive from the earliest time window, while later processing at study might not distinguish them (see Methods; Linde-Domingo et al., 2019).We split the decoded data according to which of the four *a priori* study time windows the classifiers were trained on and collapsed the results over target and non-target trials at test. In both Experiments, participants reinstated neural patterns from earlier study windows when they remembered audio versus picture items (Figure 2). This reinstatement took place predominantly in the late test window (492-776 ms), but a similar, weaker pattern was also apparent in the early test window in Experiment 2.

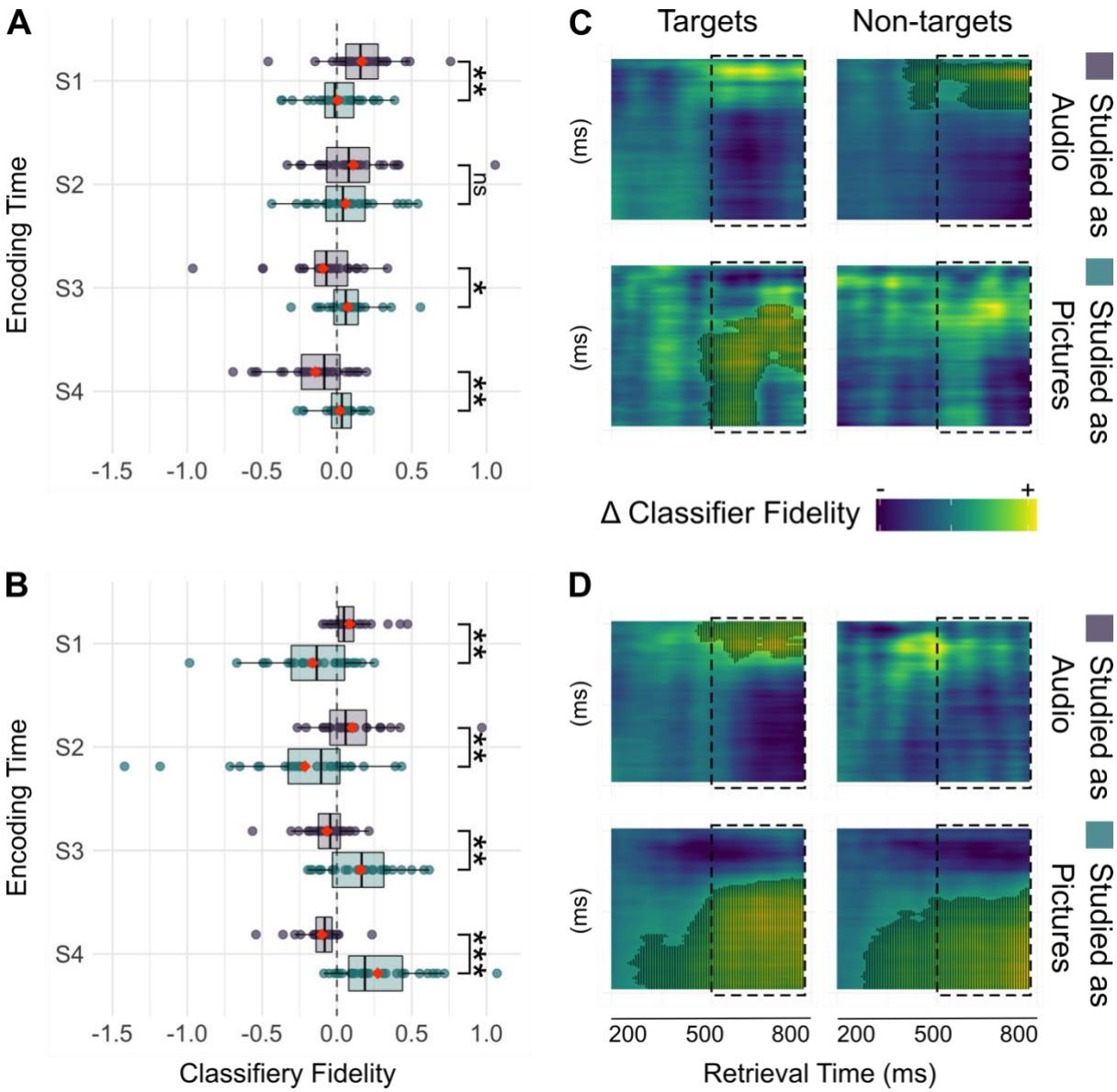

Figure 2. Format-specific reinstatement of study phase neural patterns. Upper panels show data for Experiment 1 (verbal retrieval cues) and lower panels for Experiment 2 (pictorial retrieval cues). On the left, plots show reinstatement (LDA fidelity *d*) according to study format in A) Experiment 1 and B) Experiment 2 for classifiers trained on the 4 study time

windows (S1: 136-232 ms, S2: 240-360 ms, S2: 368-672 ms, and S4: (680-984 ms) and tested on the late test time window (492-776 ms). Data points represent participants, red points are means, and boxplots show the medians and interquartile ranges. Asterisks indicate significance of tests of group-level differences (two-sided Wilcoxon signed-rank test; *p*-values are corrected with Benjamin-Hocheberg method for multiple comparisons across study-windows, i.e., * $p \leq .05$. ** $p \leq .01$. *** $p \leq .001$). On the right, images illustrate reinstatement strength after the retrieval cue, according to study phase timepoints (y axis) and test phase timepoints (x axis) in C) Experiment 1 and D) Experiment 2. The dashed areas indicate the late test time-window (492-776 ms). Reinstatement is indicated by above-zero classifier fidelity, shown in yellow. Timepoints belonging to significant clusters are highlighted with superimposed black circles. Each decoding analysis compared targets or non-targets to a baseline of new, unstudied items from the same test block (see Multivariate decoding analyses).

For Experiment 1, repeated measures ANOVA on mean classifier fidelity values with factors of study format (picture/ audio), study window (S1/ S2/ S3/ S4) and test window (early/ late) revealed a three-way interaction of study format with study window and test window, $F(1.69, 45.60) = 14.52$, $p < .001$, $\eta^2_p = 0.350$. *Post hoc* ANOVA with factors of study format and study window confirmed that reinstatement of study phase neural patterns was not significant in the early test time window, $F(1.44, 38.77) = 0.18$, $p = .766$, $\eta^2_p = .006$. In contrast, reinstatement of study neural patterns was format-specific in the late test window, with a significant interaction of study format with study time window, $F(2.30, 62.09) = 13.12$, $p < .001$, $\eta^2_p = .327$. Pairwise comparisons confirmed that neural patterns from studied pictures and audio items represented information from different study phase windows (Figure 2A and 2C). Reinstated information from the S1 window was stronger when auditory words had been studied, $p = .008$, $r = .534$, while reinstatement from the two later windows (S3 and S4) was stronger when pictures had been studied, $p = .015$, $r = .473$ and $p = .008$, $r = .555$, respectively (for S2 time window, $p = .399$, $r = .164$). We then tested when these format-specific effects were maximal. Classifier fidelity for the auditory condition was significantly greater than 0 for the S1 window only, $p < .001$, $r = .697$ (for later windows, $p = .050$, $r = .370$; $p = .997$, $r = .327$, and $p = .997$, $r = .512$). For the picture condition, reinstatement was only significant from the S3 window, $p = .029$, $r = .456$ (for S1, $p = .496$, $r = .004$; S2, $p = .196$, $r = .202$; S4, $p = .150$, $r = .270$).

In Experiment 2 a similar pattern was apparent starting from the earlier test window (Figure 2B and 2D), indicating reinstatement of study phase neural patterns prior to the onset of the left parietal ERP. ANOVA with the same factors revealed a significant three-way interaction of study format x study window x test window, $F(1.54, 41.69) = 12.55$, $p < .001$, $\eta^2_p = .317$. In the early test window *post hoc* tests showed a significant interaction of study format x study window, $F(1.29, 34.71) = 12.53$, $p < .001$, $\eta^2_p = .317$. Although reinstatement from the first two study windows was higher for audio than picture format items, and reinstatement for the last two study windows was higher for picture format items, pairwise comparisons for each study time window showed a significant difference in S4 only, $p = .006$, $r = .577$ (for S1, $p = .088$, $r = .357$; S2, $p = 0.088$, $r = .349$; S3, $p = .779$, $r = .056$).

As in Experiment 1, reinstatement was strongly format-specific in the late test window (Figure 2B). ANOVA again showed a significant interaction of study format x study window, $F(1.63, 44.06) = 24.12$, $p < .001$, $\eta^2_p = 0.472$. *Post hoc* analyses by time-window showed that neural patterns from the two early time-windows were reinstated more strongly when auditory words than pictures were remembered (for S1, $p = .002$, $r = 0.568$; for S2, $p = .009$, $r = .486$), and neural patterns from the latter two time windows were reinstated more strongly when pictures than auditory words were remembered (for S3, $p = .002$, $r = .577$; for

S4, $p < .001$, $r = .839$). To assess maximal format-specific effects, we restricted the analyses to data from the late test window. Here, classifier fidelity was significantly above 0 for auditory items from both the two early study windows and larger in S1, $p = .003$, $r = .585$; for S2, $p = 0.026$, $r = .417$ (for S3, $p = 1$, $r = .405$; S4, $p = 1$, $r = .740$). For pictures, classifier fidelity was above 0 for the later two study windows and larger in S4, $p < .001$, $r = .800$, for S3, $p = .002$, $r = .572$ (for S1, $p = .996$, $r = .495$; S2, $p = .996$, $r = .452$).

Target-selective reinstatement tracks cue overlap

Next, we tested our main experimental hypotheses by asking whether neural reinstatement was target-selective in the experimental conditions in which the left parietal ERP had shown a target-selective pattern. To compare the overall amount of target and non-target reinstatement we quantified audio and picture reinstatement from the study windows in which we had found each to be maximal: S1 for auditory reinstatement in both experiments, and S3 and S4 for picture reinstatement in Experiments 1 and 2, respectively. To examine target and non-target reinstatement when retrieval goals were held constant, we now coded items according to their target designation in the current test block rather than by whether they were studied as audios or pictures.

Figure 3 contrasts the pattern of selectivity revealed by the current reinstatement analysis with the pattern of univariate left parietal ERP selectivity previously reported by Moccia and Morcom (2021). In Experiment 1 (with verbal cues; Figure 3A), repeated measures ANOVA on mean classifier fidelity values with factors of target designation (target-picture/ target-audio), trial type (target/ non-target) and test window (early/ late) showed a significant interaction of target designation with test window, $F(1, 27) = 6.52$, p = .017, $\eta^2_p = .195$. *Post hoc* tests confirmed that reinstatement differed for targets and non-targets in the late test window only, $F(1, 27) = 4.97$, $p = .034$, $\eta^2_p = .155$ (for early window, $F(1, 27) = 0.07$, $p = .794$, $\eta^2_p = .003$). In the Target-Audio test block, memory pattern reinstatement was non-significantly greater for targets (studied as auditory words) than non-targets (studied as pictures), $p = .227$ with effect size $r = .232$. In contrast, target-selectivity was reversed in the Target-Picture block, reinstatement being stronger for *non-targets* than targets, $p = .032$, $r = .452$.

In Experiment 2 (with pictorial cues; Figure 3B) a similar but stronger and earlier-onsetting pattern of target-selectivity emerged. ANOVA with factors of target designation, trial type and test window showed a significant two-way interaction of trial type with target designation, $F(1, 27) = 7.79$, $p = .010$, $\eta^2_p = .224$, which did not vary significantly between the early and late test window $F(1, 27) = 3.96$, $p = .057$, $\eta^2_p = 128$. *Post hoc* tests confirmed target-selective neural reinstatement in the Target-Picture block, $p = .004$, with effect size $r = .568$. In the Target-Audio block, reinstatement did not significantly differ between targets and non-targets, although was numerically larger for non-targets, $p = .099$, $r = .314$.

From the above findings, it appeared that reinstatement could either be target- or non-target-selective depending on which source was targeted. We therefore wanted to know whether this pattern followed the previously reported pattern for the parietal ERP, which depended on the degree of overlap between the retrieval cues and the currently targeted source (Moccia & Morcom, 2021). We collapsed the data across experiments ($N = 56$), coding the trials according to cue-target overlap: when cues were words in Experiment 1, trials in the Target-Audio block were the high overlap condition, and when cues were pictures in Experiment 2, those in the Target-Picture block were the high overlap condition. In the late test window, neural pattern reinstatement was more target-selective in the high than the low overlap condition, $p = .001$, $r = .463$ (Figure 3C-D). *Post hoc* tests confirmed that target neural patterns were reinstated more strongly than non-target patterns in the high

overlap condition, p = .002, r = .440, while non-target neural patterns were reinstated more strongly than target patterns in the low overlap condition, p = .024, r = 0.303.

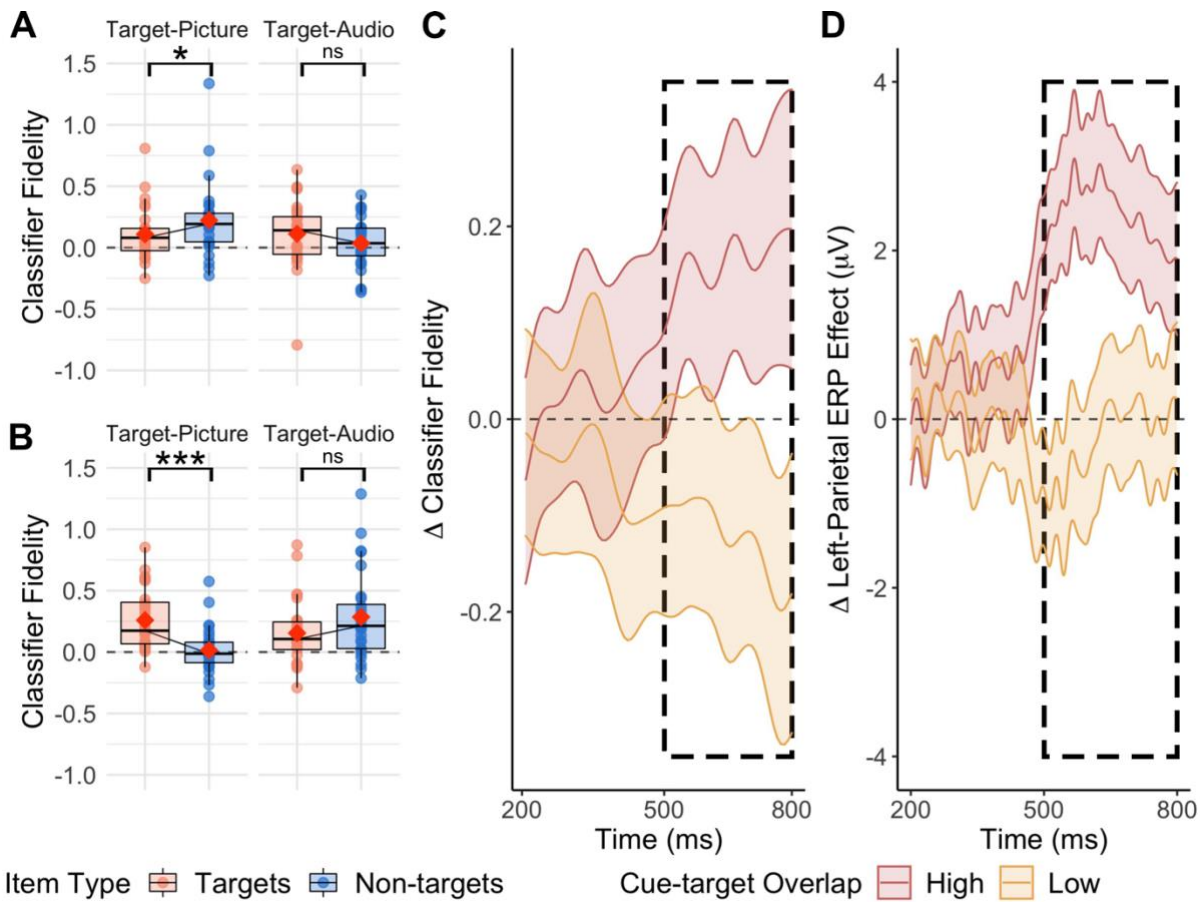

Figure 3. Target-selectivity for study-test reinstatement compared to the left parietal ERP. On the left, plots show effects of target designation – whether targets are items studied as audios or pictures – on reinstatement (LDA fidelity *d*) for targets and non-targets in A) Experiment 1 (verbal retrieval cues) and B) Experiment 2 (pictorial retrieval cues). Data points represent participants, red points are means, and boxplots show the medians and interquartile ranges. Asterisks indicate significance of tests of group-level differences (two-sided Wilcoxon signed-rank tests) in late test time window (492-776 ms); i.e., * $p \leq 0.05$. ** $p \leq 0.01$. *** $p \leq 0.001$. Panel C) shows the difference in reinstatement for targets and non-targets according to high versus low retrieval cue-target overlap, collapsed across experiments and plotted per 8 ms retrieval time bin. Panel D) shows the corresponding univariate left parietal ERP data from Moccia & Morcom (2021). The shaded areas in the line plots are the 95% within-subject confidence intervals for each time-point adjusted using the Cosineau-Morey method (Morey, 2008; cf. Craddock, 2016). The dashed rectangle indicates the analyzed time-window. For details of measures see main text.

Decoding retrieval goals from preparatory cue EEG

The foregoing analyses showed the effects of strategic control on selective retrieval by revealing selection prior to 500 ms post-stimulus. In the final set of analyses we examined neural reinstatement associated with maintaining a retrieval goal *before* any specific retrieval cue was presented. To do this, we assessed the similarity of neural patterns between the study phase and the preparatory cue interval at test (Figure 4). In Experiment 1, the LDA classifier trained to differentiate picture and audio trials at study was able to distinguish at test between

preparation to retrieve picture targets and preparation to retrieve audio targets: the classifier fidelity measure revealed three significant clusters of reinstatement from approx. 400 - 1200 ms after the preparatory cue (cluster p = .0060, .0060), and a further cluster just before retrieval cue onset (p = .023). This shows sustained proactive reinstatement of the currently relevant study context while participants prepared to target items studied in one of the two formats. However, in Experiment 2, the classifier did not significantly decode the retrieval goal (cluster p = .498).

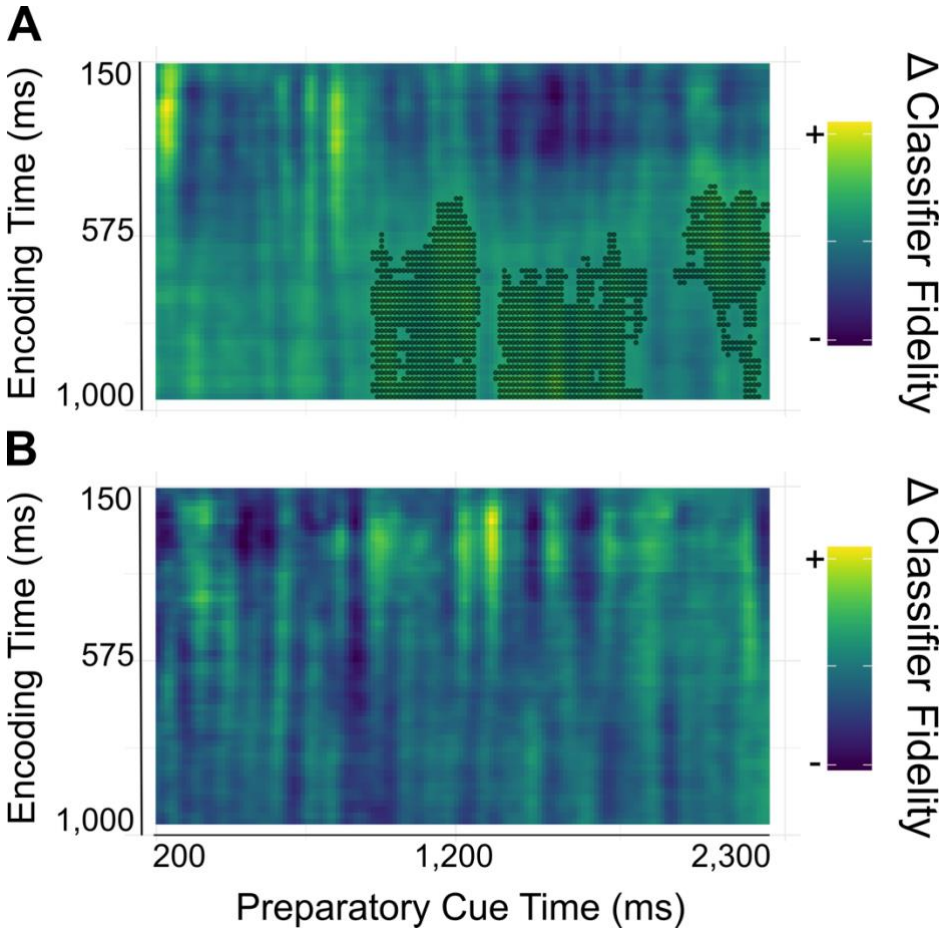

Figure 4. Goal-driven reinstatement prior to the retrieval cue. The images show the strength of reinstatement during the preparatory cue periods in A) Experiment 1 and B) Experiment 2, according to study phase timepoints (y axis) and test phase timepoints (x axis). Reinstatement is indicated by above-zero classifier fidelity, shown in yellow. Timepoints belonging to the max significant cluster are highlighted with superimposed black circles. No significant clusters were found in Experiment 2.

Discussion

A central function of memory is to recall past experiences that are relevant to current goals. Here, we used multivariate EEG decoding to investigate how and when relevant memories are selected during the retrieval cascade. We measured reinstatement of study phase neural patterns both during and before recollection in a task where goals and external cues were manipulated (Moccia & Morcom, 2021). During recollection, people reinstated temporally distinct neural patterns that reflected whether auditory words or pictures were remembered. We showed that these reinstated neural patterns were selective for targeted items when retrieval cues had high overlap with targets – i.e., when targets were auditory

words and test cues were written words (Experiment 1) or targets were pictures and test cues were line drawings (Experiment 2). The left parietal ERP was previously shown to be target-selective in these conditions (Moccia & Morcom, 2021). In contrast, when retrieval cues had higher overlap with non-targets than targets, we observed a reversed pattern with greater reinstatement of non-target information. Since the left parietal effect had been shown to be non-selective in these conditions, the two measures now diverged. These distinct patterns of selectivity revealed by the two electrophysiological measures of retrieval success support the notion that selection takes place at multiple stages before and during recollection.

Neural models of recollection specify that cortical reinstatement is triggered by hippocampal pattern completion (Alvarez & Squire, 1994; Norman & O'Reilly, 2003; Treves & Rolls, 1994). Pattern completion itself is assumed to be an automatic autoassociative process, but memory theories so far say little about the information flow during subsequent cortical memory reconstruction. For instance, it is unclear if initial pattern-completed representations are sufficient for recollection, as has been proposed (Damasio, 1989). Alternatively, reinstated information may require further, integrative and constructive processing for conscious recollection to occur, perhaps involving the posterior parietal cortex (Bergström et al., 2013; Humphreys et al., 2021; Hutchinson et al., 2009; Linde-Domingo et al., 2019; Vilberg & Rugg, 2009). On this basis, our simple model assumed that goal-directed selection might act at one or both of these two retrieval stages. If selection operates bottom-up, at the point of hippocampal pattern completion, then all subsequent activity in the retrieval cascade should be target-selective, including both reinstatement and the left parietal ERP. While we observed this pattern to some degree in two experimental conditions, the left parietal ERP in these conditions was more strongly target-selective than pattern reinstatement. Moreover, in the other two conditions, non-targeted mnemonic information was reinstated more strongly than targeted information, while the left parietal ERP was merely non-selective, with no significant difference between targets and non-targets. The data point to top-down amplification of target relative to non-target signals, with reinstated non-target information excluded from further processing.

What drives these two stages of retrieval selection? One key factor appears to be external retrieval cues. In the current study we found that neural reinstatement was cue-dependent, prioritising memories with high cue overlap regardless of whether these were currently targeted. This supports the assumption that the overlap between external partial cues and stored memory traces influences bottom-up selection by modifying pattern completion (Norman & O'Reilly, 2003). Our detection of reactivation of neural activity associated with non-targeted memories is also consistent with previous findings from tasks in which retrieval cues are associated with multiple stored memories (e.g., Bramão et al., 2022; Khader et al., 2005; Kuhl et al., 2011). The mechanisms of top-down amplification are less clear. Our data suggest that amplification cannot enable selective recollection of targeted events on its own, since the left parietal ERP was nonselective in the low cue overlap conditions which showed stronger non-target reinstatement. However, these experiments only compared two extreme conditions in which *either* targets or non-targets strongly overlapped with external cues. Goal-driven selection in the low target overlap conditions may also have been underestimated to the degree that some participants used recall-to-reject strategies that prioritised non-target retrieval (see Moccia & Morcom, 2021). Further data are needed to clarify the degree to which goal-driven top-down amplification can generate selective retrieval signals when neither target nor non-target recall is strongly cued externally, or when non-target recall is more strongly cued.

Our findings extend fMRI evidence that retrieval can be selective for targeted over non-targeted information. Both univariate activation (Elward & Rugg, 2015; Morcom & Rugg, 2012) and multivariate reinstatement can be stronger for events or event features

designated as task-relevant (Favila et al., 2018; Kuhl et al., 2013; McDuff et al., 2009). Kuhl and colleagues (2013) showed that goal-relevant information was reinstated more strongly in lateral parietal cortex, unlike in sensory cortical regions (Kuhl et al., 2013; see also Favila et al. 2018). This is compatible with the proposal that amplification involves higher-order regions like parietal cortex. However, Kuhl and colleagues (2013) proposed that the selective reinstatement in lateral parietal cortex arose from post-retrieval processes that enhance mnemonic representations in line with goals. This interpretation may be correct, but fMRI does not have sufficient time resolution to establish whether selection operates on the output of retrieval or before retrieval is completed (c.f., Rugg & Wilding, 2000). The temporal advantage of EEG allowed us to better address this theoretical question. The timing of reinstatement was broadly consistent with intracranial studies showing that cortical reinstatement follows hippocampal pattern completion from about 500 ms (Ludowig et al., 2008; Merkow et al., 2015; Mormann et al., 2005; Staresina, Fell, et al., 2012; see Staresina & Wimber, 2019 for review). The fact that reinstatement and the left parietal ERP were both target-selective between 500-800 ms post-stimulus points clearly to pre-retrieval selection. Moreover, at least in Experiment 2, reinstatement began prior to the left parietal ERP, consistent with our assumptions about their sequence within the retrieval cascade if bottom-up cue-driven processes initiate the cascade (Wimber et al., 2012). This fits with intracranial recording evidence that reinstatement takes place at different times in different brain regions (see Staresina & Wimber, 2019). However, any conclusions regarding the relative timing of selective reinstatement and amplification must be tentative, as both selective neural signals co-occurred within a broad time window. Disentangling these spatiotemporal dynamics will require methods like EEG-fMRI fusion or intracranial EEG that can measure time-resolved brain activity from specific locations.

Reinstated neural patterns in the current study differed according to whether remembered items had been studied in an auditory or picture format, corresponding to earlier or later study phase activity, respectively. Thus scalp EEG patterns depended on memory contents, converging with fMRI studies showing auditory and visual reactivation in format-consistent brain regions during retrieval (Wheeler et al., 2000; Danker & Anderson, 2010 for review). Our data add temporal specificity, indicating that information from distinct temporal stages of the study trials was reinstated when participants retrieved auditory and pictorial stimuli. Linde-Domingo et al. (2019) previously showed that reinstatement of information from earlier and later study phase time windows corresponded to perceptual and semantic information about studied pictures. Initially we hypothesised that participants might focus on semantic information when remembering pictures but emphasize phonological information when targeting auditory words (see Hornberger et al., 2004, 2006). However, although the earlier two study phase time windows of interest were (respectively) based on the timings of perceptual and semantic processing in Linde-Domingo et al.'s (2019) study, these windows did not correspond to the difference between our two study formats. This is perhaps unsurprising as the stimuli and task were different in the two studies. Although our data do not show which features of the study trial participants were reinstating, our measure of target-selectivity took these timings into account.

We also examined the control processes that may enable selective retrieval by generating internal cue representations that overlap sought-for memories (Anderson & Bjork, 1994; Burgess & Shallice, 1996; Rugg & Wilding, 2000). In Experiment 1, the classifier trained to differentiate between seeing pictures and hearing words could distinguish at test between preparing to target items in one of the two formats (Figure 4). This shows sustained, goal-related reinstatement of study context in preparation for retrieval. These results converge with earlier evidence that people engaged visual and auditory sensory cortical areas when attempting to remember studied pictures and auditory words (Hornberger et al., 2006; see

also Hornberger et al, 2004; McDuff et al., 2009). They also converge with previous studies that have shown study context reinstatement during memory search at the start of free recall (Kragel et al., 2021; Manning et al., 2011; Polyn, 2005), or successfully decoded retrieval goals prior to retrieval (Kerrén et al., 2021). In contrast, we did not detect preparatory reinstatement in Experiment 2. Although this null finding does not rule out such processing, it is possible that proactive control processing was attenuated in this task because the line-drawing cues overlapped even more completely with targeted pictures than the verbal cues did with targeted auditory words in Experiment 1, leading participants to rely more on these external cues. This possibility fits with the smaller goal-related ERPs during retrieval attempts we previously found in Experiment 2 (see Moccia & Morcom, 2021). Indirect converging evidence from scene context cueing tasks suggests that people engage less internal control when study context is externally cued at test (Smith & Vela, 2001). However, this proposal requires further, direct evaluation. Future studies with a greater number of trials will also be needed to test the additional encoding specificity prediction that preparatory reinstatement will predict successful performance on the upcoming trials. Nonetheless, the current data provide the first demonstration that preparation to retrieve elicits reinstatement of goal-relevant information, prior to a retrieval attempt.

These experiments provide several novel insights into how goals shape recollection as it occurs. We show for the first time that pre-retrieval selection can lead to selective neural reinstatement of study patterns at the time of recollection for currently targeted memories. Our data also support the proposal that selection can act at multiple stages during the retrieval processing cascade. They suggest that external cues act to modulate initial neural reinstatement bottom-up, while internal goals amplify top-down processing of prioritised memories. We also detected proactive, goal-directed reinstatement of study context while participants were preparing to retrieve targets, revealing the mental reinstatement that has been hypothesised to precede selective recollection.

Reference

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: A simple network model. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(15), 7041–7045. https://doi.org/10.1073/pnas.91.15.7041

Anderson, M. C., & Bjork, R. A. (1994). Mechanisms of inhibition in long term memory: A New Taxoomy. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory Processes in Attention, Memory, and Language.* (pp. 265–325). Academic Press.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bergström, Z. M., Henson, R. N., Taylor, J. R., & Simons, J. S. (2013). Multimodal imaging reveals the spatiotemporal dynamics of recollection. *NeuroImage*, *68*, 141–153. https://doi.org/10.1016/j.neuroimage.2012.11.030

Bramão, I., Jiang, J., Wagner, A. D., & Johansson, M. (2022). Encoding Contexts Are Incidentally Reinstated During Competitive Retrieval and Track the Temporal Dynamics of Memory Interference. *Cerebral Cortex*, bhab529. https://doi.org/10.1093/cercor/bhab529

Bramão, I., & Johansson, M. (2018). Neural Pattern Classification Tracks Transfer-Appropriate Processing in Episodic Memory. *Eneuro*, *5*(4), ENEURO.0251-18.2018. https://doi.org/10.1523/ENEURO.0251-18.2018

Buckner, R. L., & Wheeler, M. E. (2001). The Cognitive Neuroscience of Remembering. *Nature Reviews Neuroscience*, *2*(9), 624–634.

Burgess, P. W., & Shallice, T. (1996). Confabulation and the Control of Recollection. *Memory*, *4*(4), 359–412. https://doi.org/10.1080/096582196388906

Craddock, Matt (2016, November 28). ERP Visualization: Within-subject confidence intervals. https://www.mattcraddock.com/blog/2016/11/28/erp-visualization-within-subject-confidence-intervals/

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, *33*(1–2), 25–62.

Danker, J. F., & Anderson, J. R. (2010). The Ghosts of Brain States Past: Remembering Reactivates the Brain Regions Engaged During Encoding. *Psychological Bulletin*, *136*(1), 87–102. https://doi.org/10.1037/a0017937

Elward, R. L., & Rugg, M. D. (2015). Retrieval Goal Modulates Memory for Context. *Journal of Cognitive Neuroscience*, *27*(12), 2529–2540. https://doi.org/10.1162/jocn_a_00878

Favila, S. E., Samide, R., Sweigart, S. C., & Kuhl, B. A. (2018). Parietal Representations of Stimulus Features Are Amplified during Memory Retrieval and Flexibly Aligned with Top-Down Goals. *The Journal of Neuroscience*, *38*(36), 7809–7821. https://doi.org/10.1523/JNEUROSCI.0564-18.2018

Friedman, D., & Johnson, R. (2000). Event-related potential (ERP) studies of memory encoding and retrieval: A selective review. *Microscopy Research and Technique*, *51*(1), 6–28. https://doi.org/10.1002/1097-0029(20001001)51:1<6::AID-JEMT2>3.0.CO;2-R

Guggenmos, M., Sterzer, P., & Cichy, R. M. (2018). Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. *NeuroImage*, *173*, 434–447. https://doi.org/10.1016/j.neuroimage.2018.02.044

Herron, J. E., & Evans, L. H. (2018). Preparation breeds success: Brain activity predicts remembering. *Cortex*, *106*, 1–11. https://doi.org/10.1016/j.cortex.2018.04.009

Herron, J. E., & Rugg, M. D. (2003). Retrieval orientation and the control of recollection. *Journal of Cognitive Neuroscience*, *15*(6), 843–854.

Hornberger, M., Morcom, A. M., & Rugg, M. D. (2004). Neural Correlates of Retrieval Orientation: Effects of Study–Test Similarity. *Journal of Cognitive Neuroscience*, *16*(7), 1196–1210. https://doi.org/10.1162/0898929041920450

Hornberger, M., Rugg, M. D., & Henson, R. N. A. (2006). ERP correlates of retrieval orientation: Direct versus indirect memory tasks. *Brain Research*, *1071*(1), 124–136. https://doi.org/10.1016/j.brainres.2005.11.092

Humphreys, G. F., Lambon Ralph, M. A., & Simons, J. S. (2021). A Unifying Account of Angular Gyrus Contributions to Episodic and Semantic Cognition. *Trends in Neurosciences*, *44*(6), 452–463. https://doi.org/10.1016/j.tins.2021.01.006

Hutchinson, J. B., Uncapher, M. R., & Wagner, A. D. (2009). Posterior parietal cortex and episodic retrieval: Convergent and divergent effects of attention and memory. *Learning & Memory*, *16*(6), 343–356. https://doi.org/10.1101/lm.919109

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541. https://doi.org/10.1016/0749-596X(91)90025-F

Jafarpour, A., Fuentemilla, L., Horner, A. J., Penny, W., & Duzel, E. (2014). Replay of very early encoding representations during recollection. *Journal of Neuroscience*, *34*(1), 242–248. https://doi.org/10.1523/JNEUROSCI.1865-13.2014

Johnson, J. D., Price, M. H., & Leiker, E. K. (2015). Episodic retrieval involves early and sustained effects of reactivating information from encoding. *NeuroImage*, *106*, 300–310. https://doi.org/10.1016/j.neuroimage.2014.11.013

Johnson, J. D., & Rugg, M. D. (2007). Recollection and the reinstatement of encoding-related cortical activity. *Cerebral Cortex*, *17*(11), 2507–2515. https://doi.org/10.1093/cercor/bhl156

Kerrén, C., Bramão, I., Hellerstedt, R., & Johansson, M. (2021). Strategic retrieval prevents memory interference: The temporal dynamics of retrieval orientation. *Neuropsychologia*, *154*(February). https://doi.org/10.1016/j.neuropsychologia.2021.107776

Khader, P., Burke, M., Bien, S., Ranganath, C., & Rösler, F. (2005). Content-specific activation during associative long-term memory retrieval. *NeuroImage*, *27*(4), 805–816. https://doi.org/10.1016/j.neuroimage.2005.05.006

Kosslyn, S. M. (1994). *Image and Brain: The Resolution of the Imagery Debate*. MIT Press. https://doi.org/10.1176/ajp.153.9.1222

Kragel, J. E., Ezzyat, Y., Lega, B. C., Sperling, M. R., Worrell, G. A., Gross, R. E., Jobst, B. C., Sheth, S. A., Zaghloul, K. A., Stein, J. M., & Kahana, M. J. (2021). Distinct cortical systems reinstate the content and context of episodic memories. *Nature Communications*, *12*(1). https://doi.org/10.1101/2020.08.20.248567

Kuhl, B. A., Johnson, M. K., & Chun, M. M. (2013). Dissociable neural mechanisms for goal-directed versus incidental memory reactivation. *Journal of Neuroscience*, *33*(41), 16099–16109. https://doi.org/10.1523/JNEUROSCI.0207-13.2013

Kuhl, B. A., Rissman, J., Chun, M. M., & Wagner, A. D. (2011). Fidelity of neural reactivation reveals competition between memories. *Proceedings of the National Academy of Sciences*, *108*(14), 5903–5908. https://doi.org/10.1073/pnas.1016939108

Linde-Domingo, J., Treder, M. S., Kerrén, C., & Wimber, M. (2019). Evidence that neural information flow is reversed between object perception and object reconstruction from memory. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-018-08080-2

Ludowig, E., Trautner, P., Kurthen, M., Schaller, C., Bien, C. G., Elger, C. E., & Rosburg, T. (2008). Intracranially Recorded Memory-related Potentials Reveal Higher Posterior than Anterior Hippocampal Involvement in Verbal Encoding and Retrieval. *Journal of Cognitive Neuroscience*, *20*(5), 841–851. https://doi.org/10.1162/jocn.2008.20507

Manning, J. R., Polyn, S. M., Baltuch, G. H., Litt, B., & Kahana, M. J. (2011). Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. *Proceedings of the National Academy of Sciences*, *108*(31), 12893–12897. https://doi.org/10.1073/pnas.1015174108

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

McDuff, S. G. R., Frankel, H. C., & Norman, K. A. (2009). Multivoxel pattern analysis reveals increased memory targeting and reduced use of retrieved details during single-agenda source monitoring. *Journal of Neuroscience*, *29*(2), 508–516. https://doi.org/10.1523/JNEUROSCI.3587-08.2009

Mecklinger, A. (2010). The control of long-term memory: Brain systems and cognitive processes. *Neuroscience and Biobehavioral Reviews*, *34*(7), 1055–1065. https://doi.org/10.1016/j.neubiorev.2009.11.020

Merkow, M. B., Burke, J. F., & Kahana, M. J. (2015). The human hippocampus contributes to both the recollection and familiarity components of recognition memory. *Proceedings of the National Academy of Sciences*, *112*(46), 14378–14383. https://doi.org/10.1073/pnas.1513145112

Moccia, A., & Morcom, A. M. (2021). Cue overlap supports preretrieval selection in episodic memory: ERP evidence. *Cognitive, Affective, & Behavioral Neuroscience*. https://doi.org/10.3758/s13415-021-00971-0

Morcom, A. M., & Rugg, M. D. (2012). Retrieval Orientation and the Control of Recollection: An fMRI Study. *Journal of Cognitive Neuroscience*, *24*(12), 2372–2384. https://doi.org/10.1162/jocn_a_00299

Mormann, F., Fell, J., Axmacher, N., Weber, B., Lehnertz, K., Elger, C. E., & Fernández, G. (2005). Phase/amplitude reset and theta-gamma interaction in the human medial temporal lobe during a continuous word recognition memory task. *Hippocampus*, *15*(7), 890–900. https://doi.org/10.1002/hipo.20117

Morey, R. D. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau (2005). Tutorials in Quantitative Methods for Psychology, 4(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Moscovitch, M. (2008). The hippocampus as a “stupid,” domain-specific module: Implications for theories of recent and remote memory, and of imagination. *Canadian Journal of Experimental Psychology*, *62*(1), 62–79. https://doi.org/10.1037/1196-1961.62.1.62

Murray, J. G., Howie, C. A., & Donaldson, D. I. (2015). The neural mechanism underlying recollection is sensitive to the quality of episodic memory: Event related potentials reveal a some-or-none threshold. *NeuroImage*, *120*, 298–308. https://doi.org/10.1016/j.neuroimage.2015.06.069

Norman, K. A., & O’Reilly, R. C. (2003). Modeling Hippocampal and Neocortical Contributions to Recognition Memory: A Complementary-Learning-Systems Approach. *Psychological Review*, *110*(4), 611–646. https://doi.org/10.1037/0033-295X.110.4.611

Polyn, S. M. (2005). Category-Specific Cortical Activity Precedes Retrieval During Memory Search. *Science*, *310*(5756), 1963–1966. https://doi.org/10.1126/science.1117645

Polyn, S. M., Kragel, J. E., Morton, N. W., McCluey, J. D., & Cohen, Z. D. (2012). The neural dynamics of task context in free recall. *Neuropsychologia*, *50*(4), 447–457. https://doi.org/10.1016/j.neuropsychologia.2011.08.025

Rosburg, T., & Mecklinger, A. (2017). The costs of target prioritization and the external requirements for using a recall-to-reject strategy in memory exclusion tasks: A meta-analysis. *Psychonomic Bulletin and Review*, *24*(6), 1844–1855. https://doi.org/10.3758/s13423-017-1256-1

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11*(6), 251–257. https://doi.org/10.1016/j.tics.2007.04.004

Rugg, M. D., & Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Current Opinion in Neurobiology*, *23*(2), 255–260. https://doi.org/10.1016/j.conb.2012.11.005

Rugg, M. D., & Wilding, E. L. (2000). Retrieval processing and episodic memory. *Trends in Cognitive Sciences*, *4*(3), 108–115. https://doi.org/10.1016/S1364-6613(00)01445-5

Singmann H, Bolker B, Westfall J, Aust F, Ben-Shachar M (2022). _afex: Analysis of Factorial Experiments_. R package version 1.1-1,<https://CRAN.R-project.org/package=afex>.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*(2), 203–220. https://doi.org/10.3758/BF03196157

Staresina, B. P., Fell, J., Do Lam, A. T. A., Axmacher, N., & Henson, R. N. (2012). Memory signals are temporally dissociated in and across human hippocampus and perirhinal cortex. *Nature Neuroscience*, *15*(8), 1167–1173. https://doi.org/10.1038/nn.3154

Staresina, B. P., Henson, R. N. A., Kriegeskorte, N., & Alink, A. (2012). Episodic reinstatement in the medial temporal lobe. *Journal of Neuroscience*, *32*(50), 18150–18156. https://doi.org/10.1523/JNEUROSCI.4156-12.2012

Staresina, B. P., & Wimber, M. (2019). A Neural Chronometry of Memory Recall. *Trends in Cognitive Sciences*, *23*(12), 1071–1085. https://doi.org/10.1016/j.tics.2019.09.011

Thakral, P. P., Wang, T. H., & Rugg, M. D. (2015). Cortical reinstatement and the confidence and accuracy of source memory. *NeuroImage*, *109*, 118–129. https://doi.org/10.1016/j.neuroimage.2015.01.003

Thakral, P. P., Wang, T. H., & Rugg, M. D. (2017). Decoding the content of recollection within the core recollection network and beyond. *Cortex*, *91*, 101–113. https://doi.org/10.1016/j.cortex.2016.12.011

Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, *4*(3), 374–391. https://doi.org/10.1002/hipo.450040319

Tulving, E. (1983). *Elements of Episodic Memory*. Clarendon Press.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373. https://doi.org/10.1037/h0020071

Vilberg, K. L., & Rugg, M. D. (2007). Dissociation of the neural correlates of recognition memory according to familiarity, recollection, and amount of recollected information. *Neuropsychologia*, *45*(10), 2216–2225. https://doi.org/10.1016/j.neuropsychologia.2007.02.027

Vilberg, K. L., & Rugg, M. D. (2009). Functional significance of retrieval-related activity in lateral parietal cortex: Evidence from fMRI and ERPs. *Human Brain Mapping*, *30*(5), 1490–1501. https://doi.org/10.1002/hbm.20618

Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proceedings of the National*

Academy of Sciences of the United States of America, *97*(20), 11125–11129. https://doi.org/10.1073/pnas.97.20.11125

Wimber, M., Alink, A., Charest, I., Kriegeskorte, N., & Anderson, M. C. (2015). Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. *Nature Neuroscience*, *18*(4), 582–589. https://doi.org/10.1038/nn.3973

Wimber, M., Maaß, A., Staudigl, T., Richardson-Klavehn, A., & Hanslmayr, S. (2012). Rapid memory reactivation revealed by oscillatory entrainment. *Current Biology*, *22*(16), 1482–1486. https://doi.org/10.1016/j.cub.2012.05.054

Supplementary Materials

Supplementary behavioural results

Exclusion task performance. Participants' recognition performance in the exclusion task was assessed by coding items according to their target designation. These were targets (items studied as pictures in the Target-Pictures condition or as auditory words in the Target-Audio condition), non-targets and new items. Full performance is reported in Table S1 (see Moccia & Morcom, 2021 for statistical analyses).

	Verbal test cues (Experiment 1)			Pictorial test cues (Experiment 2)		
	Target	Non-target	New	Target	Non-target	New
Target-Audio block						
Accuracy	.78	.90	.87	.71	.94	.81
	[.74, .81]	[.88, .93]	[.84, .91]	[.67, .76]	[.92, .96]	[0.76, .86]
RTs	1200	1257	1264	1315	1188	1529
	[1139, 1261]	[1222, 1293]	[1199, 1,330]	[1263, 1367]	[1136, 1240]	[1467, 1592]
Target-Picture block						
Accuracy	.81	.86	.92	.89	.88	.91
	[.78, .85]	[.83, .90]	[.89, .94]	[.86, .91]	[.85, .91]	[.89, .93]
RTs	1097	1308	1294	1062	1297	1310
	[1028, 1165]	[1249, 1367]	[1224, 1363]	[1008, 1116]	[1259, 1335]	[1234, 1385]

Table S1. Memory Task Performance. Table displays means and [95% confidence intervals] for accuracy proportions and median response times (RTs in ms) for correctly identified items per condition and experiment. The Cousineau-Morey method for within-participant variables was used to adjust the confidence intervals (Morey, 2008).

Supplementary LDA results: Study phase decoding

In an initial step we used linear discriminant analysis (LDA) to decode study phase neural patterns when participants saw object pictures versus heard object names. The goal of this analysis was to inform selection of time windows for the study-test decoding between 0 and 1000 ms post stimulus. We trained the LDA classifier to discriminate between picture and audio trials by iteratively training on all trials but one and testing on the left-out trial (a leave-one-out procedure; Linde-Domingo et al., 2019). Classifier features were ERP amplitudes from the 64 scalp electrodes, after subtraction of a 200 ms baseline. Downsampling to approx. 8 ms time bins, temporal smoothing and multivariate noise normalization were applied as in the main study-test analysis. Trials were sub-sampled and mean classifier fidelity value determined over 12 iterations (see Methods, Multivariate decoding analyses in main manuscript for details).

We found that the LDA classifier could confidently assign study phase neural patterns to picture and auditory trials throughout almost the entire study phase time window (Figure S1). One-sample *t*-tests on the classifier fidelity values in each encoding time-bin showed significant above zero decoding from ~ 56 ms until ~ 1,000 ms after stimulus onset after correction for multiple comparisons (all $p \leq .001$, with exception of time-bin four: 32-40 ms, $p \leq .05$, and time-bin seven: 56-64 ms, $p \leq .01$, corrected with the Benjamini-Hochberg method; Benjamini & Hochberg, 1995). We therefore retained the two *a priori* windows in which Linde-Domingo et al. (2019) found (early) perceptual and (later) conceptual reinstatement of picture information using word cues. We then derived two later windows by dividing the subsequent time bins into two equal length windows, giving our four study windows S1 (136-232 ms), S2 (240-360 ms), S3 (368-672 ms), and S4 (680-984 ms), rounded to fit our time bins.

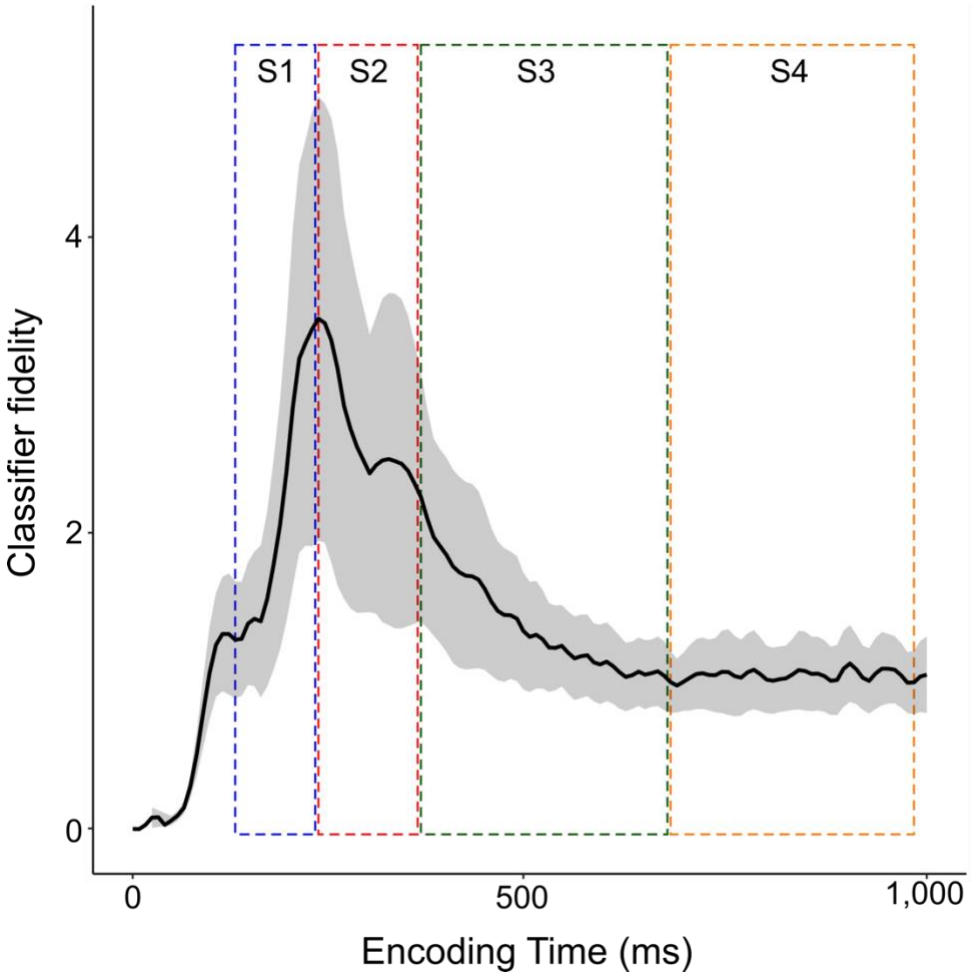

Figure S1. Mean classifier fidelity of study phase neural patterns. Coloured dashed lines are the four a priori time windows (S1: 136-232 ms, S2: 240-360 ms, S3: 368-672 ms, and S4: 680-984 ms) used for later analyses. The shaded area represents the 95% confidence intervals around the encoding time-bins showing significant above-zero decoding fidelity (see text for details).

Supplementary LDA results: Test phase decoding

Decoding neural patterns during successful retrieval. To assess whether reinstatement was present in each target and non-target condition, we decoded neural patterns elicited by the retrieval cues (see Figure 2B and 2D in main manuscript). In initial exploratory analyses, we

used LDA classifiers to test whether reinstatement of study phase neural patterns could be detected for correctly identified targets and non-targets relative to new items in the same block (see Multivariate decoding analyses in main manuscript for details on how the decoding analysis was performed). Across both experiments, this revealed significant neural reinstatement of studied information during test in approximately the same time window post-stimulus as the left parietal event-related potential (ERP). This showed that non-target as well as target reinstatement was prominent in the two conditions in which the left parietal ERP had not differed between targets and non-targets (a non-selective pattern). In the Target-Picture block in Experiment 1 (verbal test cues), we found significant clusters for both targets and non-targets (cluster $p = .018$ for targets and .035 for non-targets), and the same in the Target-Audio block of Experiment 2 (pictorial test cues; $p = .044$ and $p < .001$, respectively). However, the separate decoding of target and non-target patterns did not allow us to test whether reinstatement was target-selective in the other two conditions, which had shown target-selective left parietal ERPs. In the Target-Audio block of Experiment 1 neither target nor non-target reinstatement was separately significant ($p = .197$ and .329), while in the Target-Picture block of Experiment 2, reinstatement was only significant for targets ($p < .001$; .259 for non-targets) (see Moccia and Morcom, 2021). The next two analysis steps allowed us to directly compare the amount of target and non-target reinstatement (see Reinstated neural patterns are format-specific and Target-selective reinstatement tracks cue overlap in main manuscript).

Decoding goal-related neural patterns during retrieval attempts. To assess goal-related reinstatement during the recollection time-window, we applied LDA classifiers trained on the study phase (pictures/ auditory words) to the test phase data from new item trials. Since no information should be retrieved on new trials, any difference in reinstated neural patterns can be assumed to index differences in retrieval goals (Rugg & Wilding, 2000). Here, retrieval goals had been manipulated in the Target-Picture block versus the Target-Audio block. However, in both experiments the LDA classifier was unable to reliably distinguish at test between attempting to retrieve picture targets and attempting to retrieve audio targets. In Experiment 1, the classifier fidelity measure revealed two non-significant clusters of reinstatement from approx. 400 to 800 ms after the retrieval cue (cluster $p = .202$, .422). In Experiment 2, a slightly later-onsetting cluster was found between approximately 600 to 800 ms that was not statistically significant (cluster $p = .080$).

Supplementary results: ANOVA tables

Table S2.

Repeated Measures ANOVAs: Format-specific Reinstatement Effects

	Experiment 1					Experiment 2				
Effect	df	MSE	*F*	η^2_p	*p*	df	MSE	*F*	η^2_p	*p*
Study Window	1.36, 36.84	0.06	8.14 **	.232	.004	1.39, 37.59	0.04	20.35 ***	.430	<.001
Test Window	1, 27	0.02	1.29	.046	.266	1, 27	0.04	3.49 +	.115	.072
Study Format	1, 27	0.20	0.54	.020	.469	1, 27	0.26	0.14	.005	.708
Study Window×Test Window	1.23, 33.31	0.02	10.72 **	.284	.001	1.32, 35.77	0.02	1.84	.064	.182
Study Window×Study Format	2.20, 59.43	0.04	6.21 **	.187	.003	1.56, 42.05	0.13	23.69 ***	.467	<.001
Test Window×Study Format	1, 27	0.04	0.03	.001	.869	1, 27	0.04	2.01	.069	.168
Study Window×Test Window×Study Format	1.69, 45.60	0.02	14.52 ***	.350	<.001	1.54, 41.69	0.04	12.55 ***	.317	<.001

+ $p \leq .10$, * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$.

Table S3.

Follow-up Repeated Measures ANOVAs: Format-specific Reinstatement Effects in Early Test-Window (T1)

	Experiment 1					Experiment 2				
Effect	df	MSE	*F*	η^2_p	*p*	df	MSE	*F*	η^2_p	*p*
Study Window	1.52, 41.14	0.02	1.98	.068	.160	1.30, 35.01	0.02	14.01 ***	.342	<.001
Study Format	1, 27	0.10	0.63	.023	.435	1, 27	0.11	1.02	.037	.321
Study Window×Study Format	1.44, 38.78	0.03	0.18	.006	.766	1.29, 34.71	0.06	12.53 ***	.317	<.001

+ $p \leq .10$, * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$.

Table S4.

Follow-up Repeated Measures ANOVAs: Format-specific Reinstatement Effects in Late Test-Window (T2)

	Experiment 1					Experiment 2				
Effect	df	MSE	*F*	η^2_p	*p*	df	MSE	*F*	η^2_p	*p*
Study Window	1.28, 34.48	0.06	10.86 **	.287	.001	1.43, 38.61	0.04	14.09 ***	.343	<.001
Study Format	1, 27	0.13	0.32	.012	.575	1, 27	0.19	0.03	<.001	.873
Study Window×Study Format	2.30, 62.09	0.04	13.12 ***	.327	<.001	1.63, 44.06	0.12	24.11 ***	.472	<.001

+ $p \leq .10$, * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$.

Table S5.

Repeated Measures ANOVAs: Selective Reinstatement Effects

	Experiment 1					Experiment 2				
Effect	df	MSE	*F*	η^2_p	*p*	df	MSE	*F*	η^2_p	*p*
Target Designation	1, 27	0.07	4.94 *	.155	.035	1, 27	0.07	2.92 +	.098	.099
Item Type	1, 27	0.09	0.26	.009	.615	1, 27	0.10	0.41	.015	.525
Test Window	1, 27	0.03	17.20 ***	.389	<.001	1, 27	0.04	22.19 ***	.451	<.001
Target Designation×Item Type	1, 27	0.10	0.96	.034	.335	1, 27	0.14	7.79 **	.224	.010
Target Designation×Test Window	1, 27	0.01	0.90	.032	.350	1, 27	0.01	1.87	.065	.183
Item Type×Test Window	1, 27	0.01	0.02	<.001	.899	1, 27	0.01	3.84 +	.124	.061
Target Designation×Item Type×Test Window	1, 27	0.02	6.52 *	.195	.017	1, 27	0.03	3.96 +	.128	.057

+ $p \leq .10$, * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$.

Table S6.

Follow-up Repeated Measures ANOVAs: Selective Reinstatement Effects in Early Test-Window (T1).

Experiment 1					
Effect	df	MSE	*F*	η^2_p	*p*
Target Designation	1, 27	0.03	3.66 +	.119	.066
Item Type	1, 27	0.05	0.29	.011	.596
Target Designation×Item Type	1, 27	0.07	0.07	.003	.794

+ $p \leq .10$, * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$.

Table S7.

Follow-up Repeated Measures ANOVAs: Selective Reinstatement Effects in Late Test-Window (T2)

Experiment 1					
Effect	df	MSE	*F*	η^2_p	*p*
Target Designation	1, 27	0.05	4.59 *	.145	.041
Item Type	1, 27	0.06	0.17	.006	.680
Target Designation×Item Type	1, 27	0.05	4.97 *	.155	.034

+ $p \leq .10$, * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$.

Reference

Moccia, A., & Morcom, A. M. (2021). Cue overlap supports preretrieval selection in episodic memory: ERP evidence. *Cognitive, Affective, & Behavioral Neuroscience*. https://doi.org/10.3758/s13415-021-00971-0

Morey, R. D. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau (2005). Tutorials in Quantitative Methods for Psychology, 4(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061