Strategic retrieval prevents memory interference: The temporal dynamics of retrieval orientation

Casper Kerrén a,*, Inês Bramão a, Robin Hellerstedt b, Mikael Johansson a

a Department of Psychology, Lund University, Sweden
b School of Psychology, University of Kent, United Kingdom

A R T I C L E   I N F O
Keywords:
Retrieval orientation
Retention-induced forgetting
Episodic memory
Reinstatement
Multivariate pattern-analysis
Interference

A B S T R A C T
Resolving interference between overlapping memories is crucial to remember the past. This study tests the novel prediction that orienting search focus benefits goal-relevant retrieval by reducing competition from unwanted memories. In a modified retrieval-practice paradigm, participants encoded word-pairs in one of two encoding tasks. Critically, to evaluate whether this retrieval orientation (RO) reduces memory interference, target and competitor memories were always related to different encoding tasks. At retrieval, instructions were provided for half of the blocks with the intention to bias remembering towards items encoded with one of the ROs. Behavioural data show that adopting an RO improved target accessibility, strengthened the testing effect, and reduced retrieval-induced forgetting (RIF) of competitors. Specifically, RIF—typically attributed to inhibitory control of memory interference—was prominent when no retrieval orientation (NRO) instruction was provided. Furthermore, a neural correlate of RO was calculated by training a linear discriminant analysis (LDA) to discriminate the electroencephalographic (EEG) spatial brain patterns correspondent to the two ROs over the time course of selective retrieval. RO was characterised by increases in the theta and decreases in the beta frequency band, evident both before and after category-cue onset. While the pre-cue RO reinstatement effect predicted both immediate retrieval-practice success and later target accessibility, the post-cue effect predicted disengagement of inhibitory control, such that participants showing a stronger RO reinstatement effect showed lower levels of RIF. These data suggest that strategically orienting search focus during retrieval both increases target memory accessibility and reduces memory interference, which consequently protects related memories from inhibition and later forgetting. Furthermore, they also highlight the roles of theta and beta oscillations in establishing and maintaining a task-relevant bias towards target memory representations during competitive memory retrieval.

1. Introduction

Episodic memory allows us to mentally travel in time to relive personally experienced events tied in time and place (Tulving, 1972; Tulving and Thomson, 1973). However, access to our past is often hindered by competition arising from similar overlapping events. For example, trying to remember what you had for dinner the last time you went to a particular restaurant will likely reactivate other instances you have been to both that and probably also other restaurants. While previous research has shown that strategic retrieval processes may prepare the retrieval of certain mnemonic content (Rugg and Wilding, 2000), it remains unclear how such a retrieval orientation contributes to competitive episodic remembering. The present study uses multivariate pattern analysis (MVPA) implemented by linear discriminant analysis (LDA) of high temporal-resolution electroencephalographic (EEG) data to understand if adopting a retrieval orientation benefits selective target retrieval by reducing interference from related competing memories.

Episodic remembering occurs when a retrieval cue triggers the reinstatement of neural processes that were engaged at the time of the original event (e.g., Marr, 1971; Norman and O’Reilly, 2003). The likelihood of successful retrieval increases as a function of the overlap between the processing elicited by the retrieval cue and that engaged during encoding (Bramão and Johansson, 2018; Morris et al., 1977; Nyberg et al., 2000; Ritchey et al., 2013; Rugg et al., 2008; Staresina et al., 2012; Tulving and Thomson, 1973; Wheeler et al., 2000). Recent work has linked the neural reinstatement during episodic remembering to both synchronized and desynchronized oscillatory brain activity (e.g., the Sync/deSync model; Parish et al., 2017). Specifically, retrieval is
associated with an increase in theta power (3–7 Hz; Guderian et al., 2009; Addante et al., 2011; Fell et al., 2013; Gruber et al., 2013) and a decrease in alpha/beta power (8–30 Hz; Griffiths, Martin-Buro, Star-esina and Hanslmayr, 2020; Griffiths et al., 2019; Parish et al., 2017; Parish et al., 2020; Waldhauser et al., 2012).

Often, however, retrieval cues do not only overlap with the sought-after memory but cause the reactivation of other related memories that compete for retrieval. Both neural network modelling (Norman et al., 2006; Norman et al., 2007; Ritvo et al., 2019) and empirical studies (Favila et al., 2016; Hulbert and Norman, 2015; Kim et al., 2017; Schlichting et al., 2015) have shown that retrieval competition increases as a function of the similarity between the memories’ neural representations. The inhibitory control account holds that competitive retrieval is resolved by actively inhibiting the competing memories (Anderson, 2003; Hulbert et al., 2012; Levy and Anderson, 2002; Storm and Levy, 2012). A long-term consequence of such inhibition is decreased accessibility of the competitors, a phenomenon referred to as retrieval-induced forgetting (RIF; Anderson et al., 1994; see Anderson, 2003, Murayama et al., 2014, for reviews). However, the goal relevance associated with an increase in theta power (3–7 Hz; Guderian et al., 2009; Addante et al., 2011; Fell et al., 2013; Gruber et al., 2013) and a decrease in alpha/beta power (8–30 Hz; Griffiths, Martin-Buro, Star-esina and Hanslmayr, 2020; Griffiths et al., 2019; Parish et al., 2017; Parish et al., 2020; Waldhauser et al., 2012).

2. Materials and methods

2.1. Participants

Thirty-two right-handed participants (mean 23.3 ± 2.5 years), with normal or corrected-to-normal vision and with no reported history of neurological diseases, voluntarily took part in the experiment in exchange for a cinema ticket. The study was conducted in accordance with the Swedish Act concerning the Ethical Review of Research involving Humans (2003:460). Participants gave their written informed consent and the study followed the local ethical guidelines at Lund University.

2.2. Stimuli material and experimental design

A total of 320 nouns, comprised of 10 different words belonging to 32 different semantic categories, were selected from a Swedish category norm set (Hellerstedt et al., 2012). To generate a controlled and yet heterogeneous range of competition strength in our stimuli material, we selected words with both low and high taxonomic frequency. This created two different conditions: a low and a high target-competitor taxonomic frequency similarity. Items with high taxonomic frequency are more frequently provided as exemplars within a given semantic category (Hellerstedt et al., 2012). Thus, within each category, we selected 1) four words with low taxonomic frequency that worked as targets (e.g., Sport – Cricket; mean-rank 45 across category lists); 2) three words with low taxonomic frequency that worked as competitors in the high target-competitor similarity condition (e.g., Sport – Boule; mean-rank 44.7 across category lists); and 3) three words with high taxonomic frequency that functioned as competitors in the low target-competitor similarity condition (e.g., Sport – Golf; mean-rank 8.2 across category-lists). Importantly, following the recommendations of Anderson (2003) we excluded words with very high taxonomic frequency (rank 1–4) for each category. Thus, the high taxonomic frequency similarity condition consisted of target and competitor items from the same taxonomic frequency range, whereas the low similarity condition contained target and competitor items from different taxonomic frequency ranges. All 10 exemplars, within each category, had a unique initial letter that served as probe during retrieval.

The material, consisting of 32 categories with 10 words each, was divided into four lists of eight semantic categories each, matched by rank, F(3, 77) = 0.014, p = .93 and word length, F(3, 77) = 0.019, p = 1. The words were between 4 and 10 letters long. The four lists were counterbalanced across the conditions of interest (Retrieval Orientation (Retrieval Orientation vs. No Retrieval Orientation) & Similarity (High vs. Low)). However, since no interactions involving the Similarity factor were observed, the results are presented collapsed across this factor.

Additionally, to investigate the consequences of competitive retrieval, two out of eight categories in each list were excluded from the intermediate retrieval-practice phase and served as baseline items for
the RIF and for the testing effect.

2.3. Procedure

The experiment comprised four blocks, two in which participants received a retrieval orientation (RO) instruction and two in which no such instruction was provided (NRO). The blocks were shown in alternating order. Half of the participants started with an RO block and the other half started with an NRO block. Each block included an encoding phase, a retrieval-practice phase, a distractor phase, and a test phase.

In each block, during encoding, participants were presented with 56 exemplars (28 category-targets and 28 category-competitors associations). Participants were asked to encode all the category-exemplar associations, each presented in one of two different encoding tasks. In the artist task, participants were asked to rate on a 3-point scale (1 = difficult and 3 = easy) How easy is this (i.e., the exemplar) to paint for an artist? In the like task, participants were asked to rate on a 3-point scale (1 = not much and 3 = a lot) How much do you like this (i.e., the exemplar)? The encoding tasks were adapted from Dzulkifli and Wilding (2005) and were the basis for the retrieval orientation instruction provided in the retrieval-practice phase. Half of the exemplars were encoded with the artist task and the other half with the like task. Importantly, to narrow participants search focus, target and competitor items were always encoded with different tasks. Items were studied in random order. After an initial fixation cross (1500 ms), participants were presented with either a K! (konstnär [artist]) or a G! (gilla [like]) for 600 ms, informing participants which task to use when encoding the association. The category-exemplar (Sport-Boule) presentation followed and remained on screen for 6000 ms. After 4000 ms the rating scale was presented at the bottom of the screen, prompting participants to respond, and remained on screen until the end of the trial (see Fig. 1).

The retrieval-practice phase followed the encoding phase. Participants’ memory performance for 24 out of the 28 category-target exemplars presented during encoding was tested in a random order. This is the only phase where the RO and the NRO conditions differed. In the RO blocks, participants were instructed that the upcoming 12 items were encoded together with one of the two encoding tasks. In one block, participants started by retrieving items encoded in the artist task, and in the other RO block participants started by retrieving items encoded in the like task. In the NRO blocks no such instruction was given and the task information was provided on a trial-by-trial basis to equate the presentation of the two conditions. Items were retrieved in random order. Notice that the trial structure was identical for both conditions. An initial fixation cross (1500 ms) started the trial. After this, the category-cue and the letter corresponding to the encoding task was presented for 600 ms. Next, the first letter of the exemplar was shown for 3000 ms (Sport-B__) and participants were asked to retrieve the exemplar. Participants responded orally and their response was registered by the experimenter. To avoid muscle artifacts in the EEG data, participants were asked to withhold their response until the presentation of a question mark. Participants performed three retrieval attempts for each category-exemplar association (See Fig. 1).

To ensure that retrieval in final memory test was not attributable to working memory, participants engaged in a counting-backwards distractor task for 30 s. Following this, participants were tested on all category-exemplar associations from the encoding phase. Again, a fixation cross (1500 ms) initiated the trial. The category-cue together with the first letter of the exemplar was shown for 3000 ms (Sport-B__). When prompted with an interrogation mark, participants had 2000 ms to respond. Output interference was controlled for by testing competitor exemplars prior to target exemplars (Anderson et al., 1994; Roediger, 1974).

The stimuli were presented on a 17-inch LCD display at 60 Hz using E-Prime in a sound-attenuated room using a Faraday cage.

2.4. EEG data acquisition and preprocessing

Preprocessing was done using the FieldTrip toolbox (Oostenveld et al., 2011) and custom written MATLAB code. The electroencephalogram (EEG) was recorded continuously at a sampling rate of 2048 Hz from 32 silver/silver chloride scalp electrodes mounted in an elastic cap and positioned according to the extended 10–20 system (Jasper, 1958). Electrooculogram electrodes (EOG) were placed over and below the left eye and at the right and left outer canthi. Online, the electrodes were referenced to the left mastoid. The impedance between the electrode and the scalp was kept below 5 kΩ during the entire experiment.

The continuous EEG data were downsampled to 512 Hz and filtered with a Butterworth high-pass filter of 1 Hz, a low-pass filter of 200 Hz, and a band-stop filter (50 Hz; 100 Hz, and 150 Hz). The EEG data corresponding to the retrieval-practice phase were then divided into trials from 2000 ms pre-stimulus to 4000 ms post-stimulus onset and referenced offline to the average of the left and right mastoid bones electrodes. A semi-automatized trial- and independent component rejection was applied to the data. In a first step, the data were high-pass filtered to 100 Hz and trials that were further away than four times the median absolute deviation from the amplitude distribution were automatically removed. In a second step, ICA was used to detect artifacts to be removed in the data. To limit computational load, we took advantage of that ICA finds regularities in the data and we therefore used the first 1.5 s segments and one third of the trials to find components to exclude. The unmixing matrix was then applied to the data. In order to remove blink artifacts, the components showing spectral and topographical patterns of blink artifacts were removed. An automated additional artefact rejection was then conducted, similar to the initial step. Again, data were low- and high-pass filtered to 100 Hz and 1 Hz, and the trials that were further away than three times the median absolute deviation from the distribution were rejected. Three times the median absolute deviation was chosen, for a more conservative approach. After components and trials were removed, all trials were again visually inspected, and trials still containing artifacts were manually removed. Bad channels were interpolated using the triangulation method as implemented in FieldTrip. On average 132 out of 144 trials were kept for the RO condition (min = 113, max = 143, SD = 7.65), and on average 133 out of 144 trials were kept for the NRO condition (min = 106, max = 142, SD
In order to investigate the reinstatement of the retrieval orientation task during the retrieval-practice phase, a Linear Discriminant Analysis (LDA) was used (Treder, 2020; https://github.com/treder/MVPA-Light). The model was trained to discriminate the two encoding tasks over the time course of the retrieval practice, using the amplitude of the spatial activation pattern from the EEG channels. The obtained weights were then used at each time point to predict which class the brain pattern belonged to, resulting in a time-generalization matrix. To speed up the computations, data were downsampled to 128 Hz before smoothed with a 100 ms running average using MATLAB’s function smoothdata. To avoid overfitting, the data were cross-validated using a leave-one-out approach, meaning that for each time point, the classifier was trained on all but one trial, and tested on the remaining trial.

The relevant output of the classifier was twofold: 1) the average decoding performance, i.e., the probability that the classifier made a correct decision based on the true label of the data, which was calculated by comparing the predicted label to the true label, for each time point; and 2) the distance value (d-value or fidelity value). The former resulted in a time series of decoding performance, in which 50% was considered chance level. To infer whether the results obtained were significantly exceeding chance level, a permutation test (as implemented in FieldTriptool) between conditions of interest was conducted. Fidelity values were used to investigate the oscillatory component of the decoding, explained below (2.7). Importantly, we were interested in the average performance across all electrodes and both measures therefore lack spatial specificity.

2.6. Functional relationship between the retrieval orientation reinstatement effects and behavioural memory performance measures

The RO reinstatement effect was calculated by subtracting the decoding performance obtained in the NRO from the RO condition. A median split based on the obtained reinstatement effect was conducted and two groups of participants were created: a high RO reinstatement group and a low RO reinstatement group. Importantly, because retrieval-induced forgetting was only reliable in the second half of the experiment, participants were split based on an RO reinstatement effect calculated only by using trials from the second half of the experiment. To investigate if the high reinstatement group showed stronger benefits of the retrieval orientation instruction, both in terms of target accessibility and competitor forgetting, we contrasted the memory performance measures in the RO against the NRO condition in the two groups. The RO behavioural index was calculated by subtracting the behavioural memory performance in the NRO condition from the behavioural memory performance in the RO condition.

2.7. The oscillatory components of significant decoding

The fidelity values show the distance to the hyperplane that can best separate the two classes of the decoding (in this case the two encoding tasks) and have previously been shown to have a functional relationship with episodic memory reinstatement (Kerren et al., 2018; Linde-Domingo et al., 2019). We followed the procedure in previous studies (Kerren et al., 2018; Linde-Domingo et al., 2019) and used the MODAL algorithm (Watrous et al., 2017) to estimate the power of the fidelity values of the significant RO reinstatement effects. Importantly, the MODAL algorithm searches for the presence of oscillations across time points by investigating the change in phase per unit time. Briefly, the obtained fidelity values in each sample were reshaped into one time series and z-scored before subjected to a fast-fourier transformation (FFT). A plateau-shaped bandpass filter was constructed using the lowest and highest frequencies obtained in the FFT exceeding the fitted line in log-log space using robustfit in MATLAB (Lega et al., 2012) and was applied to the raw data. Next, the Hilbert transform was applied to the filtered data to obtain the analytic signal and the phase angle time series was then extracted. Frequency and phase at each sample was obtained using a frequency sliding method. Next, phase slips were attenuated by applying median filters with different length in the time domain (50 ms–400 ms), wherefrom the median was taken (Cohen, 2014). Importantly, only frequencies that exceeded the 1/f-fitted line were included, resulting in a vector for each participant containing the frequencies and time points where an oscillation was present. We then extracted power values at time points in which an oscillation was observed. To evaluate whether the power values obtained exceeded what was expected by chance, they were contrasted with a baseline. To this end, the same classification as previously described was run, but now using random labels for training the classifier. This was repeated 25 times. We then applied the MODAL algorithm to the 25 random label classifiers and took the 2.5th and 97.5th percentile of the estimated chance distribution. If the power values in the real data exceeded those in the randomly shuffled data, they were deemed significant. Importantly, only participants (21) in which the algorithm found an oscillation were included in the statistical tests.

2.8. Event-related potential (ERP) analysis

Several studies have shown that participants experience more interference in the first round compared to the following rounds of retrieval-practice (Kuhl et al., 2007; Rafidi et al., 2018; Wimber et al., 2015). Accordingly, we calculated ERPs for the first and the third round of the retrieval-practice separately for the RO and for the NRO conditions. The data were low-pass filtered to 30 Hz and baseline corrected from −1000 to −500 ms. The time window of interest was obtained by first collapsing the conditions (RO and NRO) and calculating the ERP between the first and third round. This is an unbiased way of deciding on a time window of interest (Cohen, 2014). To evaluate whether there was a statistical difference between the first and the third round collapsed across conditions, a non-parametric cluster-based permutation test was conducted. The permutation was run 1000 times on data from −1000 ms to 4000 ms post-stimulus onset, and the minimum number of neighbouring channels that were considered an effect was set to two. Only positive effects were considered, such that the first round of retrieval-practice had a more positive-going ERP compared to the third round. We then performed four paired-sample t-tests for each condition within the time windows for which the permutation procedure identified a significant effect, contrasting first and third round, using the Fz-electrode. Fz was chosen based on previous findings of frontal-midline (FM) EEG/ERP interference effects in episodic memory (Hellerstedt and Johansson, 2014; Hellerstedt et al., 2016; Hsieh and Ranganath, 2014; Waldhauser et al., 2012).

3. Results

3.1. Memory performance

3.1.1. Retrieval-practice phase

Cued-recall accuracy, quantified as the percentage of correct responses, was calculated for each participant and condition. We predicted that the benefits of adopting an RO would be more evident when participants would have had the opportunity to learn how to strategically make use of the RO instructions (second half of the experiment). To investigate this, a 2 by 2 by 3 repeated-measures ANOVA with the factors Retrieval Orientation (RO vs. NRO), Half (First vs. Second) and Retrieval Rounds (Round 1 vs. Round 2 vs. Round 3) was conducted. The analysis revealed main effects of Retrieval Orientation, F(1,31) = 5.89, p = .021, Half, F(1,31) = 26.26, p < .001, and Retrieval Rounds, F(2,30) = 54.71, p < .001. As predicted, subjects were better at retrieving the target memories in the RO condition (mean ± SD = 53.08 ± 8.05%).
compared with the NRO condition (mean ± SD = 49.05 ± 9.60%; see Fig. 2a). Additionally, participants retrieved significantly more items in the second half of the experiment (mean ± SD = 55.64 ± 9.83%) compared with the first half (mean ± SD = 46.48% ± 8.18%) and memory performance increased in each round (Round 1: mean ± SD = 48.21 ± 6.56%; Round 2: mean ± SD = 50.85 ± 8.45%; Round 3: mean ± SD = 54.13 ± 8.16%). However, no interactions between the factors were observed, suggesting comparable benefits of adopting an RO throughout the experiment in terms of target retrieval.

### 3.1.2. Final memory test

In the final test phase, participants were tested on all category-exemplar associations from the encoding phase. First, we predicted to observe a stronger testing effect for the RO compared with the NRO condition. Second, forgetting of competing memories (i.e., RIF) was hypothesized to be reduced when adopting an RO. Again, we predicted that these effects would be more evident in the second half of the experiment.

We investigated the testing effect by contrasting memory performance for targets that belonged to categories that were included in the intermediate retrieval-practice phase with memory performance for targets that belonged to categories that were excluded from the intermediate retrieval-practice phase. Significance level was set to 0.0125 (0.05/4) to control for multiple comparisons involving four conditions. Not surprisingly, participants’ memory performance in the final test was better for items that were practiced in the intermediate retrieval phase. Reliable testing effects were observed in all four conditions of interest: RO first half, \( t(31) = 6.27, p < .001 \); NRO first half, \( t(31) = 3.18, p = .003 \); RO second half, \( t(31) = 7.16, p < .001 \); NRO second half, \( t(31) = 5.75, p < .001 \). Next, we investigated if the testing effect was stronger in the RO condition and more so in the second half of the experiment. We conducted a 2 by 2 repeated-measures ANOVA with the factors Retrieval Orientation (RO vs. NRO) and Half (First vs. Second). As predicted, the testing effect was stronger in the RO compared with the NRO condition (main effect of Retrieval Orientation, \( F(1,31) = 4.19, p = .049 \)). However, no interaction between the two effects were observed which indicated that the RO-induced benefit is comparable across the first and the second half of the experiment (Fig. 2b).

Next, we investigated how RO modulates forgetting of competing memories. We contrasted memory performance for competitors that belonged to categories that were part of the intermediate retrieval phase with memory performance for competitors that belonged to categories that were excluded from the intermediate retrieval phase. A RIF-effect is evident when memory performance in the final test is reduced for competitors of practiced categories compared with memory performance for exemplars from unpractised categories. Again, the significance level was set to 0.0125 (0.05/4) to control for the number of statistical tests computed. The results revealed a significant RIF effect only for the NRO condition in the second half of the experiment, \( t(31) = -3.31, p = .002 \). The RIF effect for the RO in the first half of the experiment did not survive correction for multiple comparisons, \( t(31) = -2.05, p = .048 \) and no reliable RIF was observed for the other two conditions: NRO first half, \( t(31) = -1.16, p = .25 \); RO second half, \( t(31) = -1.40, p = .17 \).

Subsequently, we investigated whether participants experienced less interference and thus less RIF as a function of time in the experiment. A 2 by 2 repeated-measures ANOVA with the factors Retrieval Orientation (RO vs. NRO) and Half (First vs. Second) was performed. The analysis revealed a significant interaction between the two factors, \( F(1,31) = 15.38, p < .001 \). Follow up \( t \)-tests revealed significant differences between RO and NRO in second half of experiment, \( t(31) = 2.41, p = .02 \), but not in first half, \( t(31) = -1.23, p = .23 \). These findings suggest that adopting an RO reduces interference and limits the need for inhibitory control, and furthermore that participants may need some time on this demanding retrieval-practice task to efficiently engage cognitive control to handle interference from competing memories (cf. Román, Soriano, Gómez-Arziza and Bajo, 2008).

Together these data suggest that adopting an RO enhances both immediate and later target retrieval and at the same time limits forgetting of the competing material. We next turn to the electrophysiological data to elucidate the potential neural mechanism mediating the benefits of a retrieval orientation strategy on memory accessibility and memory interference.

### 3.2. Retrieval orientation reinstatement effects

To investigate the reinstatement of a retrieval orientation, we trained classifiers to distinguish the amplitude of the spatial EEG patterns that belonged to items that were encoded in the two different encoding tasks. The reinstatement of the encoding task, over the time-course of the
retrieval-practice, was used as a proxy for how much participants engaged in strategic retrieval and was done separately for the RO and NRO conditions. The classifiers were trained and tested on each time point during retrieval practice, which resulted in a time-generalization matrix. The time-generalization matrix allowed us to investigate if the retrieval orientation strategy was temporally widespread over the course of the retrieval-practice epoch.

Two significant reinstatement effects were observed in the RO condition: an early effect in the pre-cue time window spanning from −1000 to 500 ms, and a late effect starting 2000 ms post-category cue onset. For the NRO condition, no significant reinstatement was found (Fig. 3a).

When contrasting the two conditions, two effects showed stronger reinstatement in the RO compared with the NRO condition. The first effect started −1000 ms before category-cue onset and extended until onset. The second effect started around 2000 ms after category-cue onset and lasted until the end of trial (Fig. 3b). The time-generalization matrix clearly shows that the effects span both vertically and horizontally off the diagonal, indicating that RO is a cognitive state spanning across time. However, notice that there is no overlap between the two effects, which suggests that they are two different aspects of the RO state.

To further understand the relationship between the two RO reinstatement effects, we performed a correlation. The analysis revealed a significant positive correlation ($r = .58, p = .004$), indicating that participants showing a strong reinstatement effect in the early time window also showed a strong reinstatement effect in the late time window. Moreover, to visualize the contribution of the different spatial features of each effect, the weights of each feature were plotted separately for the two different RO reinstatement effects. As can be seen in Fig. 3c, the topographies of these effects are widespread, but there is an apparent stronger contribution of the right hemisphere in the late effect.

In sum, these results indicate that brain patterns of the encoding task were reinstated in the RO condition, when participants engaged in retrieval orientation strategies. Two RO effects were observed: an early preparatory pre-cue effect and a late post-cue effect. Even though the two effects are related and show somewhat similar topographies, the time-generalization matrix clearly shows that these effects represent two different aspects of an RO state.

### 3.3. Functional relationship between the retrieval orientation reinstatement effect and memory performance

To investigate the functional relationship between the RO reinstatement effects and episodic memory performance, we did a median-split based on the RO reinstatement effect, resulting in two groups of participants (i.e., a high and a low RO reinstatement group) for each of the two RO reinstatement effects (i.e., pre-cue and post-cue). Subsequently, the behavioural memory performance between the RO and the NRO conditions were contrasted in the two groups. The high reinstatement group in the pre-cue time window showed a significant behavioural benefit of retrieving the target in the RO condition compared with the NRO condition, which was evident both in the retrieval-practice phase (high reinstatement group, $t(1,15) = 3.60, p = .003$; low reinstatement group, $t(1,15) = 0.17, p = .87$; see Fig. 4a), and in the final memory test (high reinstatement group, $t(1,15) = 4.92, p = .0001$; low reinstatement group, $t(1,15) = 0.09, p = .93$; see Fig. 4b). This indicates that higher RO reinstatement effect in the pre-cue time window is

![Fig. 3. Decoding results.](image-url)
associated with stronger behavioural benefits of strategic target retrieval. However, the post-cue RO reinstatement effect did not conclusively explain the benefits of strategic target retrieval (retrieval-practice phase: high reinstatement group: t(1,15) = 1.87, p = .08; low reinstatement group: t(1,15) = 1.50, p = .15; final test phase: high reinstatement group: t(1,15) = 2.14, p = .049; low reinstatement group: t(1,15) = 1.7, p = .11).

Next, we investigated the functional relationship between the RO reinstatement effects and forgetting of the competitors. Comparable RIF in the RO and NRO condition were observed between the high and low reinstatement group considering the pre-cue RO reinstatement effect (high reinstatement group: t(1,15) = 1.56, p = .14; low reinstatement group: t(1,15) = 1.78, p = .09). More interestingly, the post-cue RO reinstatement effect was associated with RIF. Participants in the low reinstatement group displayed comparable levels of RIF in the RO and NRO condition, t(1,15) = 0.78, p = .45, whereas the high reinstatement group showed reduced RIF in the RO condition, t(1,15) = 2.69, p = .017); see Fig. 4c.

In sum, our results show that the RO reinstatement effects have a relationship with the behavioural measures of episodic retrieval. While the pre-cue reinstatement effect predicts immediate and future target memory accessibility, the post-cue reinstatement effect predicts the reduced forgetting of the competing memory traces when adopting a RO.

3.4. Oscillatory components of the retrieval orientation reinstatement effect

We estimated the oscillatory component of the fidelity values of the diagonal of the two retrieval orientation effects using the MODAL algorithm (Watrous et al., 2017). We observed that both RO reinstatement effects were associated with an increase in theta power (4–7 Hz) and a decrease in beta power (13–30 Hz) compared to what was expected by chance (see Fig. 3d).

In a subsample of participants, in which we identified significant theta/beta oscillations (N = 21), we investigated if the magnitude of the theta increase and beta decrease differed between the two RO reinstatement effects. A 2 by 2 repeated-measures ANOVA with the factors Frequency Band (Theta vs. Beta) and Effect (Pre vs. Post) was performed. The ANOVA revealed a borderline interaction effect (Frequency band × Effect interaction), F(1,20) = 3.48, p = .077. An exploratory analysis showed that the beta power was different between the two RO reinstatement effects, t(1,20) = 6.03, p < .001, such that the beta power decrease was more prominent in the post effect (mean = −11.97, SD = 3.83) as compared to the pre (mean = −5.57, SD = 3.24). However, the theta power was not significantly different between the two effects, t(1,20) = 2.10, p = .049 (pre: mean = 4.92, SD = 5.78; post: mean = 1.83, SD = 2.39), with a statistical threshold set to 0.025 to control for multiple tests.

In sum, the RO reinstatement effects, measured using power of the fidelity values, were associated with an increase in theta power and a decrease in beta frequency band, extending the literature on the role of theta and beta oscillations in strategic retrieval. Additionally, we observed that the early pre-cue RO reinstatement effect was associated with a numerical increase in theta power, while the late post-cue RO reinstatement effect was associated with a further decrease in beta power. This is in line with previous studies showing that memory retrieval is associated with an increase in theta in the pre-retrieval time window and with a decrease in alpha/beta during the post-retrieval time window (Addante et al., 2011; Hanslmayr et al., 2012; Mizrak et al., 2018).

3.5. Neural ERP correlates of memory interference control

As a final step, we analysed the electrophysiological correlates of memory competition. Previous studies have shown that retrieval competition and interference control are associated with positive-going modulations over frontal regions (e.g., Hellerstedt and Johansson, 2014; Hellerstedt et al., 2016). In the present study, behavioural RIF was found for the NRO condition in the second half of the experiment, suggesting that inhibitory control was engaged to a larger extent. Therefore, we tested whether the previously reported ERP correlate of memory
interference was more evident in the NRO compared with the RO condition, and more so in the second half of the experiment. An unbiased time window of interest was obtained by collapsing trials across all conditions (Cohen, 2014) and then contrasting first and third round of retrieval practice, in line with previous studies investigating competitive memory retrieval (Kuhl et al., 2007; Rafidi et al., 2018; Waldhauser et al., 2012; Wimber et al., 2015). This revealed a significant effect over frontal electrodes from approximately 2400 to 3100 ms after cue onset. We used this time window to test for differences within each condition. Again, the statistical threshold was set to 0.0125 to control for multiple tests. Results revealed a memory interference effect at Fz for the NRO condition in both the first, $t(1,31) = 2.80, p = .009$ and the second half, $t (1,31) = 3.62, p = .001$, of the experiment. Critically, this was not the case for the RO condition, first half: $t(1,31) = 1.94, p = .06$; second half: $t (1,31) = 0.82, p = .42$ (Fig. 5).

To investigate the functional relevance of this putative neural correlate of interference in the NRO condition, we used the same rationale as in 3.3. We sorted the ERP differences between first and third round and did a median split based on the amplitude difference, which resulted in a high ERP interference group and in a low ERP interference group. Next, we investigated the behavioural RIF effect in these two groups of participants. The statistical test was set to .025 (0.05/2) to correct for multiple tests. Critically, only the high ERP interference group showed a significant RIF: high, $t(1,15) = −3.12, p = .007$; low, $t (1,15) = −2.27, p = .038$ (Fig. 5c), indicating that a higher ERP interference effect tracked the higher engagement of inhibitory control.

4. Discussion

Adopting different retrieval orientations (ROs) alters the way retrieval cues are processed based on the goals of the retrieval task (Rugg and Wilding, 2000). Although previous studies have shown a functional role of adopting an RO on target accessibility (e.g., Bridger et al., 2009), no previous study has investigated the role of RO in competitive memory retrieval. In the present study, we instructed participants to orient search focus to target-specific memory representations, in an adapted retrieval-practice paradigm. Our results show that when participants were instructed to set up an RO towards goal-relevant mnemonic content, they remembered more target memories and, at the same time, showed reduced retrieval-induced forgetting (i.e., less forgetting of competing memories). A neural pattern classifier was trained to distinguish the reinstatement of the two encoding tasks during the retrieval-practice phase. This revealed two different RO reinstatement effects: a pre-cue effect that predicted both immediate and future target retrieval and a post-cue effect that predicted reduced competitor forgetting. Both RO reinstatement effects oscillated at theta and beta frequency.

The behavioural results are consistent with predictions. Adopting an RO not only improves immediate access during retrieval practice, in line with previous findings (Bridger et al., 2009; Bridger and Mecklinger, 2012; Evans and Herron, 2019; Roberts et al., 2014; Wilckens et al., 2012), it also boosts later memory performance for the practiced targets; that is, strengthening the so-called testing effect (Roediger and Butler 2011, Fig. 2a–b). Several memory theories have offered explanations to the increased target memory performance following repeated retrieval. For example, the elaborative processing view (Glover, 1989; Whitten and Leonard, 1980) holds that memory performance on the final memory test will depend on how elaborate the intervening retrieval phase is. In the present study, the additional information provided in the RO condition might have allowed participants to elaborate more in the retrieval-practice phase and thereby incorporate the RO into the memory trace. In the NRO condition, the relevant encoding task of the target
was signalled on a trial-by-trial basis and was therefore difficult to utilise. This fits well with the idea that the retrieval period allows the activated memory representation to be integrated with stored neocortical knowledge (Anthony et al., 2017).

Along with increased memory performance for target memories, the present behavioural results also revealed reduced RIF in the RO condition, while significant RIF only in the NRO condition (Fig. 2c). This supports the hypothesis that adopting an RO enhances target retrieval without the expense of forgetting the competing material. Our findings can be interpreted in light of the recent non-monotonic plasticity hypothesis (NMHP) account (Ritvo et al., 2019). This account explains how the neural representation of competitor and target memories changes as a function of competitor reactivation during retrieval. If, during retrieval, the related memories are moderately active they will interfere with target memories, and lead to the memory representations becoming less similar and competing memories being inhibited. On the other hand, if competing memories have low activation, they will not interfere with target memories and no change will take place (Ritvo et al., 2019). Thus, a possible consequence of a pre-retrieval orientation towards goal-relevant target memories is an increase in the relative difference in target and competitor activation during retrieval. According to the NMHP, this may have resulted in a low activation level of the competitors (leading to less interference) and no relative neural change for the competitor memory representations (leading to less forgetting of the competitors). On the other hand, when participants were not instructed to orient retrieval towards target memories, both targets and competitor memory traces are reactivated and compete for retrieval. Consequently, control mechanisms (Anderson, 2003) handle the interference by active inhibition of the competing memories, leading to the observed typical RIF-effect in the NRO condition.

In the present study, we examined RO with a novel experimental design in which we compared a blocked to an interleaved condition. Our hypothesis was that participants could only adopt an RO in the blocked condition. A pattern classifier was trained to distinguish the reinstated neural patterns of two different encoding tasks during retrieval practice. The assumption was that an RO state would involve the reinstatement of the neural activity associated with encoding contextual details characterizing the goal-relevant target memories (i.e., the encoding task; cf. Dzulkifli and Wilding, 2005). Our results confirmed this idea and we observed both a pre-cue and a post-cue reinstatement effect, likely related with initiating and maintaining an RO, respectively (Rugg and Wilding, 2000, Fig. 3a-c). This aligns with previous ERP studies and extends this literature by showing pre-cue and a post-cue RO effects in a novel paradigm (Bridger et al., 2009; Dzulkifli et al., 2004; Dzulkifli and Wilding, 2005; Herron & Rugg, 2003a, 2003b; Ranganath and Paller, 1999; Rosburg et al., 2013; Rosburg et al., 2014; Rosburg, Mecklinger, & Johansson, 2011a, 2011b; Rugg and Wilding, 2000; Stenberg et al., 2006). The pre-cue RO effect started 1000 ms before cue onset and continued off-the-diagonal to approximately 500 ms after cue onset. The second effect had an onset approximately 2000 ms after cue onset, and lasted, both on- and off-the-diagonal, until the response had been made. Thus, the observed RO patterns are consistent with a tonically maintained cognitive state, with the brain activity generalising to surrounding time points (King and Dehaene 2014, Fig. 5a). Moreover, a significant positive correlation between the first and second effect was observed. Previous studies have also shown a correlation between the pre- and the post-cue effect (Addante et al., 2011; Herron and Rugg, 2003a, 2003b). Critically, however, we observed different contributions of these RO reinstatement effects to memory performance. While the pre-cue effect is related to target memory, both during retrieval practice and during test (Fig. 4a-b), the post-cue effect is related to RIF (Fig. 4c). These results corroborate and extend the current literature by suggesting that adopting an RO involves at least two stages that both contribute to successful performance during competitive memory retrieval.

The observed RO reinstatement effects comprised both a theta power increase and a beta power decrease in the pre- and post-cue time windows. Previous studies have shown that both theta and alpha/beta activity in the post-cue time window are related with successful retrieval (see: Kahana et al., 2001; Herweg et al., 2020, for reviews on theta activity). At the same time a growing body of evidence shows that processes, reflected in pre-cue theta activity, are also important for later successful retrieval Addante et al. (2011); Fell et al. (2011); Gruber et al. (2013); Guderian et al. (2009); Mizrak et al. (2018); Rutishauser et al. (2010). Our results indicate that the spectral power differed between the two time windows and frequencies. More specifically, we observed a trend for the theta power to be stronger in the pre-cue time window compared to the post-cue time window. Interestingly, we observed a stronger beta power decrease in the post-cue RO reinstatement effect. Speculatively, this stronger spectral reduction in the post-cue RO reinstatement effect could reflect the need for participants to maintain the RO, search and reinstate the associated exemplar, and lastly judge whether the retrieved exemplar was correct (Burgess and Shallice, 1996; Mitchell et al., 2004). That is, participants needed to represent more information later on in the trial than early on, which would result in stronger beta desynchronization (Hanslmayr et al., 2012; Parisi et al., 2017).

Having established that adopting an RO increased memory performance for target memories and reduced forgetting of competitor memories in the RO condition, we next wanted to investigate previously reported EEG correlates of RIF (Ferreira et al., 2014; Hanslmayr et al., 2010; Hellerstedt and Johansson, 2014). Specifically, we examined whether the ERP correlate of retrieval competition and interference control was reduced in the RO condition, but present in the NRO condition. The results confirm this and show an effect only in the NRO condition, indicating that when participants could not utilise the encoding task and narrow their search focus, they suffered more interference. Interestingly, although we found behavioural RIF only in the second half of the experiment, the ERP results suggest significant interference already in the first half (Fig. 5a, right). Importantly, the more interference, as seen in more positive-going ERP in the first repetition as compared to the third repetition, the more RIF was observed in the final test phase (Fig. 5c). Although the ERP effect was functionally related to RIF it is interesting to note that the time was significantly later than in previous studies (Ferreira et al., 2014; Hanslmayr et al., 2010; Hellerstedt and Johansson, 2014). In the present study, we instructed participants to withhold the response after having received the category-exemplar cue because we wanted clean EEG data. This could potentially have postponed or sustained the interference effect. However, in both Hellerstedt and Johansson (2014) and Rafidi et al. (2018), participants were told to withhold their response and these studies still reported an early interference effect (~400 ms post-cue). A second possibility is that the list length and greater complexity from adding contextual information (i.e., the source memory information) made this novel paradigm substantially more difficult as compared to the commonly used retrieval-practice paradigm. Consequently, increased processing time and longer time to search for the relevant memory would be expected.

5. Conclusion

The present study investigated how memory control processes such as ROs can facilitate memory performance. Complementing the approaches in previous studies, we used multivariate pattern analysis to identify RO neural reinstatement effects that predicted enhanced target retrieval and limited forgetting of competing memories. Only when participants could not strategically focus on the relevant mnemonic content (NRO condition), we observed neural correlates of interference. Thus, competitive retrieval may not always lead to forgetting. Together, the results provide new knowledge of the importance of mnemonic control mechanisms for successful episodic remembering.
neocortical alpha/beta oscillations in human episodic memory. bioRxiv 305698. https://doi.org/10.1101/305698.

Gruber, M.J., Watrous, A.J., Ekstrom, A.D., Ranganath, C., Otten, I.J. 2013. Expected reward modulates encoding-related theta activity before an event. Neuroimage 64, 68-74. https://doi.org/10.1016/j.neuroimage.2012.07.064.

Guderman, S., Schott, B.H., Richardson-Klavehn, A., Duzel, E. 2009. Medial temporal theta before an event: evidence for episodic encoding in humans. Proc. Natl. Acad. Sci. U. S. A. 106 (13), 5365-5370. https://doi.org/10.1073/pnas.0900289106.

Hanslmayr, S., Staudigl, T., Aslan, A., Baum, K.H. 2010. Theta oscillations predict the detrimental effects of memory retrieval. Cognit. Affect. Behav. Neurosci. 10 (3), 329-338. https://doi.org/10.3758/CABN.10.3.329.

Hanslmayr, S., Staudigl, T., Fellner, M.C., 2012. Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. Front. Hum. Neurosci. 6, 74. https://doi.org/10.3389/fnhum.2012.00074.

Hellerstedt, R., Rasmussen, A., Johansson, M., 2012. Swedish Category Norms. Department of Psychology, Lund University, Lund, Sweden.

Hellerstedt, R., Johansson, M., 2014. Electrophysiological correlates of competitor activation predict retrieval-induced forgetting. Cerebr. Cortex 24 (6), 1619-1629. https://doi.org/10.1093/cercor/bht015.

Hellerstedt, R., Johansson, M., Anderson, M.C., 2016. Tracking the intrusion of unwanted memories into awareness with event-related potentials. Neurobiology 89, 510-523. https://doi.org/10.1016/j.neurobiology.2016.07.008.

Herron, J.E., Rugg, M.D., 2003a. Retrieval orientation and the control of recollection. J. Cognit. Neurosci. 15 (6), 843-854. https://doi.org/10.1162/089892903769404415.

Herron, J.E., Rugg, M.D., 2003b. Strategic influences on recollection: evidence for electrophysiological dissociation. Psychol. Bull. Rev. 10 (3), 703-710. https://doi.org/10.3758/BF03196535.

Herron, J.E., Wilding, E.L., 2004a. An electrophysiological dissociation of retrieval mode and retrieval orientation. Neuroimage 22 (4), 1554-1562. https://doi.org/10.1016/j.neuroimage.2004.04.011.

Herron, J.E., Wilding, E.L., 2006. Neural correlates of control processes engaged before and during recovery of information from episodic memory. Neuroimage 30 (2), 634-644. https://doi.org/10.1016/j.neuroimage.2005.10.003.

Herweg, N.A., Solomon, E.A., Kahana, M.J., 2020. Theta oscillations in human memory. Trends Cognit. Sci. 24 (3), 208-227. https://doi.org/10.1016/j.tics.2019.12.006.

Hornberger, M., Morcom, A.M., Rugg, M.D., 2004. Neural correlates of retrieval orientation: effects of study-test similarity. J. Cognit. Neurosci. 16 (7), 1196-1210. 2004.

Hornberger, M., Rugg, M.D., Heaton, R.N., 2006. ERP correlates of retrieval orientation: direct versus indirect memory tasks. Brain Res. 1071 (1), 124-136. https://doi.org/10.1016/j.brainres.2005.02.048.

Hsieh, L.T., Ranganath, C., 2014. Frontal midline theta oscillations during working memory maintenance and episodic encoding and retrieval. Neuroimage 85 (Pt 2), 721-729. https://doi.org/10.1016/j.neuroimage.2013.08.003.

Hultert, J.C., Norman, K.A., 2015. Neural differentiation tracks improved recall of unwanted memories following interleaved study and retrieval practice. Cerebr. Cortex 25 (10), 3994-4008. https://doi.org/10.1093/cercor/bhz284.

Hultert, J.C., Shidlev, G., Anderson, M.C., 2012. Evidence against associative blocking as a cause of cue-independent retrieval-induced forgetting. Exp. Psychol. 59 (1), 11-21. https://doi.org/10.1016/j.expbpsych.2011.09.005.

Jasper, H.H. 1958. The ten-twenty electrode system of the international federation. Electroencephalogr. Clin. Neurophysiol. 10, 371-375.

Kahana, M.J., Seelig, D., Madsen, J.R., 2001. Theta returns. Curr. Opin. Neurobiol. 11 (6), 739-744. Retrieved from. https://www.ncbi.nlm.nih.gov/pubmed/11741027.

Keren, C., Linde-Domingo, J., Hanslmayr, S., Wimber, M., 2018. An optimal oscillatory phase for pattern reactivation during memory retrieval. Curr. Biol. 28 (21), 3383-3392. https://doi.org/10.1016/j.cub.2018.08.065 e3386.

Kim, G., Norman, K.A., Turk-Browne, N.B., 2017. Neural differentiation tracks improved recall of unexpected memories. J. Neurosci. 37 (8), 2022-2031. https://doi.org/10.1523/JNEUROSCI.0251-18.2018.

King, J.R., Dehaene, S., 2014. Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cognit. Sci. 18 (4), 203-210. https://doi.org/10.1016/j.tics.2014.01.002.

Kuhl, B.A., Dudakov, N.M., Kahn, I., Wagner, A.D., 2007. Decreased demands on cognitive control reveal the neural processing benefits of forgetting. Nat. Neurosci. 10 (7), 908-914. https://doi.org/10.1038/nn1918.

Lega, B.C., Jacobs, J., Kahana, M., 2012. Human hippocampal theta oscillations and the formation of episodic memories. Hippocampus 22 (4), 748-761. https://doi.org/10.1002/hipo.20937.

Levy, B.J., Anderson, M.C., 2002. Inhibitory processes and the control of memory retrieval. Trends Cognit. Sci. 6 (7), 299-305. https://doi.org/10.1016/s1364-6613(02)01953-3.

Linde-Domingo, J., Tredar, M.S., Keren, C., Wimber, M., 2019. Evidence that neural information flow is reversed between object perception and object reconstruction from memory. Nat. Commun. 10 (1), 179. https://doi.org/10.1038/s41467-018-0929-2.

Marr, D., 1971. Simple memory: a theory for archicortex. Philos. Trans. R. Soc. Lond. B Biol. Sci. 262 (841), 23-81. Retrieved from. https://www.ncbi.nlm.nih.gov/pubmed/18914412.

Mitchell, K., Johnson, M., Raye, C., Greene, E., 2004. Prefrontal activity associated with source monitoring in a working memory task. J. Cognit. Neurosci. 16 (6), 921-934. https://doi.org/10.1162/0898929042646602.

Mizrahi, E., Kim, K., Roberts, B., Ragland, D.J., Carter, C., Ranganath, C., 2018. Impact of oscillatory α/θ disturbances on hippocampal gamma activity during source memory retrieval. Neuroimage 9 (1-3), 194-207. https://doi.org/10.1016/j.neuroimage.2018.12.073.
