

Top-Down Task Goals Induce the Retrieval State

Devyn E. Smith and  Nicole M. Long

Department of Psychology, University of Virginia, Charlottesville, VA 22904

Engaging the retrieval state (Tulving, 1983) impacts processing and behavior (Long and Kuhl, 2019, 2021; Smith et al., 2022), but the extent to which top-down factors—explicit instructions and goals—versus bottom-up factors—stimulus properties such as repetition and similarity—jointly or independently induce the retrieval state is unclear. Identifying the impact of bottom-up and top-down factors on retrieval state engagement is critical for understanding how control of task-relevant versus task-irrelevant brain states influence cognition. We conducted between-subjects recognition memory tasks on male and female human participants in which we varied test phase goals. We recorded scalp electroencephalography and used an independently validated mnemonic state classifier (Long, 2023) to measure retrieval state engagement as a function of top-down task goals (recognize old vs detect new items) and bottom-up stimulus repetition (hits vs correct rejections (CRs)). We find that whereas the retrieval state is engaged for hits regardless of top-down goals, the retrieval state is only engaged during CRs when the top-down goal is to recognize old items. Furthermore, retrieval state engagement is greater for low compared to high confidence hits when the task goal is to recognize old items. Together, these results suggest that top-down demands to recognize old items induce the retrieval state independent from bottom-up factors, potentially reflecting the recruitment of internal attention to enable access of a stored representation.

Key words: brain state; EEG; encoding; retrieval; MVPA

Significance Statement

Both top-down goals and automatic bottom-up influences may lead us into a retrieval brain state—a whole-brain pattern of activity that supports our ability to remember the past. Here we tested the extent to which top-down versus bottom-up factors independently influence the retrieval state by manipulating participants' goals and stimulus repetition during a memory test. We find that in response to the top-down goal to recognize old items, the retrieval state is engaged for both old and new probes, suggesting that top-down and bottom-up factors independently engage the retrieval state. Our interpretation is that top-down demands recruit internal attention in service of the attempt to access a stored representation.

Introduction

Despite evidence that a tonically maintained retrieval state (or mode; Tulving, 1983) impacts behavior (Long and Kuhl, 2019), the factors that govern how the retrieval state is engaged are unclear. Although considered a goal-driven, intentional state (Rugg and Wilding, 2000), the retrieval state may also be induced automatically (Duncan et al., 2012; Smith et al., 2022). If asked if you have seen the film *Best in Show*, you may turn your mind's eye inward and conclude that you have not. The top-down demand (try to retrieve *Best in Show*) may lead you to intentionally engage the retrieval state. Alternatively, after hearing the title, you might automatically bring to mind images of Parker Posey

melting down in a pet shop; here, bottom-up signals—activation of a stored representation—automatically pull you into the retrieval state. As these factors are typically considered in isolation, the relative contribution of bottom-up and top-down factors to retrieval state engagement remains an open question. An additional challenge is that the exact nature of the retrieval brain state—and its relation to attentional states—remains unclear. Addressing these questions is essential to our understanding, as top-down versus bottom-up driven states likely recruit distinct control mechanisms. The aim of this study was to identify the joint impact of bottom-up and top-down factors on retrieval state engagement.

The retrieval state may constitute an intentional state that is a precursor to successful retrieval (Lepage et al., 2000; Herron and Wilding, 2004, 2006) or may be induced automatically by bottom-up factors (Duncan et al., 2012; Smith et al., 2022). An individual may explicitly engage a brain state—a collection of whole-brain activity/connectivity patterns sustained over time (Harris and Thiele, 2011)—when given the goal to retrieve a stored representation. Alternatively, perceiving an item as old may induce a lingering retrieval state (Duncan et al., 2012).

Received March 1, 2024; revised May 28, 2024; accepted June 17, 2024.

Author Contributions: D.E.S. and N.M.L. designed research; D.E.S. and N.M.L. performed research; D.E.S. and N.M.L. analyzed data; D.E.S. and N.M.L. wrote the first draft of the paper; D.E.S. and N.M.L. edited the paper; D.E.S. and N.M.L. wrote the paper.

This work was supported by a grant from the National Institutes of Health (NINDS R01 NS132872, PI: N.M.L.).

The authors declare no competing financial interests.

Correspondence should be addressed to Nicole M. Long at nlong@virginia.edu.

<https://doi.org/10.1523/JNEUROSCI.0452-24.2024>

Copyright © 2024 the authors

Using an explicit mnemonic state task in combination with multivariate decoding methods, we have shown that individuals both flexibly engage the retrieval state in response to top-down demands (Long and Kuhl, 2019) and enter a retrieval state automatically when stimuli have overlapping temporal contexts (Smith et al., 2022). Together, these findings suggest a role for both bottom-up and top-down factors in the induction of the retrieval state.

Regardless of its induction, an important open question is the extent to which the retrieval state reflects internal attention. Attention can be divided along multiple dimensions (Corbetta and Shulman, 2002; Nobre, 2018), including an internal/external axis (Chun et al., 2011), whereby a focus on sensory, environmental information constitutes external attention and a focus on thoughts, representations, and task sets constitutes internal attention. According to this framework, episodic memory retrieval may fall under the internal attention axis (Long et al., 2018; Tarder-Stoll et al., 2020) and recent computational modeling and neural work have provided evidence for this account (Logan et al., 2021; Long, 2023).

Insofar as the retrieval state reflects internally directed attention, bottom-up versus top-down factors should differentially modulate when and how the retrieval state is engaged. Specifically, top-down demands should recruit internal attention *prior* to representation access. That is, the retrieval state should be engaged in response to “Have you seen Best in Show” before you have located the representation in your mind. Difficult-to-access representations should recruit the retrieval state to subserve memory search and evidence accumulation (Polyn and Kahana, 2008; Ratcliff and McKoon, 2008). Critically, top-down-driven retrieval should emerge whenever there are explicit retrieval demands. In contrast, bottom-up induction of the retrieval state should occur *after* representation access. After automatic activation of Parker Posey, your mind’s eye focuses inward, away from the external world. Easy-to-access representations should recruit the retrieval state via reflexive attentional capture (Cabeza et al., 2011).

Our hypothesis is that top-down and bottom-up factors do not interact and specifically that top-down demands fully engage the retrieval state. Alternatively, both factors may be additive such that top-down demands coupled with stimulus repetition produce a stronger retrieval state than either factor alone. To test our hypothesis, we conducted independent recognition memory studies in which we varied top-down test phase goals while recording scalp electroencephalography (EEG). We used cross-study classification to measure test phase retrieval state engagement. Because the classifier was trained to distinguish memory states, we refer to “retrieval state evidence” throughout the text; however, we expect that this signal broadly reflects internal attention demands. We predicted that retrieval state evidence would increase for all test trials when task demands required participants to recognize old stimuli and stored representations were difficult to access.

Materials and Methods

Participants

Seventy-six native English speakers from the University of Virginia community participated, with 38 participants enrolled in each experiment (E1: 28 female; age range = 18–32, mean age = 20.47 years; E2: 26 female; age range = 18–32, mean age = 20.5 years). All participants had normal or corrected-to-normal vision. Informed consent was obtained in accordance with University of Virginia Institutional Review Board for Social and Behavioral Research and participants were compensated for their participation. Data collection for the two experiments was interleaved

to prevent systematic biases between experiments that might otherwise arise due to temporal (day/week/month/semester) differences when the data were collected. Our sample size was determined *a priori* based on behavioral pilot data (E1, $N=5$; E2, $N=3$) described in the pre-registration report of this study (<https://osf.io/6u9px>). A total of four participants (two each from E1 and E2) were excluded from the final dataset due to EEG event markers not being recorded. Thus data are reported for the remaining 72 participants.

Experimental design

We conducted two recognition memory experiments (E1 and E2) and manipulated test phase instructions between subjects. Participants’ goal was to successfully retrieve study items (E1) or to detect new items (E2). Stimuli consisted of 1,602 words, drawn from the Toronto Noun Pool (Friendly et al., 1982). From this set, 640 words were randomly selected for each participant. Participants completed four phases (Fig. 1). Phase 3 was divided into two subsets; the practice subset preceded phase 1 and the main subset preceded phase 3.

Phase 1. In each of 10 runs, participants viewed a list containing 16 words, yielding a total of 160 trials. On each trial, participants saw a single word presented for 2,000 ms followed by a 1,000 ms inter-stimulus interval (ISI). In E1, participants were instructed to study the presented word in anticipation for a later memory test and did not make any behavioral responses. In E2, participants were instructed to read the words silently and did not make any behavioral responses. We have reproduced the phase 1 instructions below (omitting instructions regarding sitting still, taking breaks, for brevity).

E1 phase 1 instructions. During this experiment, you will be studying lists of words. Your task is to try to remember the words for a memory test.... In each run you will view individual words and have 2 seconds to study each word. There will be fixation crosses (plus signs) shown after each item. You won’t make a response on the keyboard, we just want you to look at the word, think about it, and try to remember it.

E2 phase 1 instructions. During this experiment, you will be evaluating words. In this first phase, your task is to read each of the words silently to yourself. You’ll see these words again in another phase of the experiment.... In each run you will view individual words and have 2 seconds to read each word. There will be fixation crosses (plus signs) shown after each item. You won’t make a response on the keyboard, we just want you to look at the word, read it, and think about it.

Phase 2. Participants completed a recognition memory test with different memory goals. On each trial, participants viewed a word which had either been presented during phase 1 (target) or had not been presented (lure). In E1, participants’ task was to make an old or new judgment for each word using a confidence rating scale from 1 to 4, with 1 being definitely new and 4 being definitely old. In E2, the task was framed as a detection phase in which participants’ task was to detect new words that were not presented in phase 1. Participants made an old or new judgment without the use of a confidence rating scale for each word by pressing one of two buttons (“d” or “k”). Response mappings were counterbalanced across participants. Phase 2 trials were self-paced and separated by a 1,000 ms ISI. There were a total of 320 test trials with all 160 phase 1 words presented along with 160 novel lures. We have reproduced the phase 2 instructions below (omitting instructions regarding sitting still, taking breaks, for brevity).

E1 phase 2 instructions. During the test phase, you will see some of the words you studied along with new words that you haven’t seen before. On each trial, you will be presented with a word. The word could be one you’ve seen before, or it could be an entirely new word. We would like you to rank how confident you are that each word is new or old on a scale from 1 to 4.

E2 phase 2 instructions. Now you are going to perform a detection task. You’ll again see words. Some words you will have seen during the reading phase, some words you will not have seen during the reading phase. Your goal is to detect all of the new words that you haven’t seen before. You should press D/K whenever you detect a new word. You should press K/D to reject any words that aren’t new. It’s really important that you detect all of the new words, but don’t stress if you forget some of the old words.

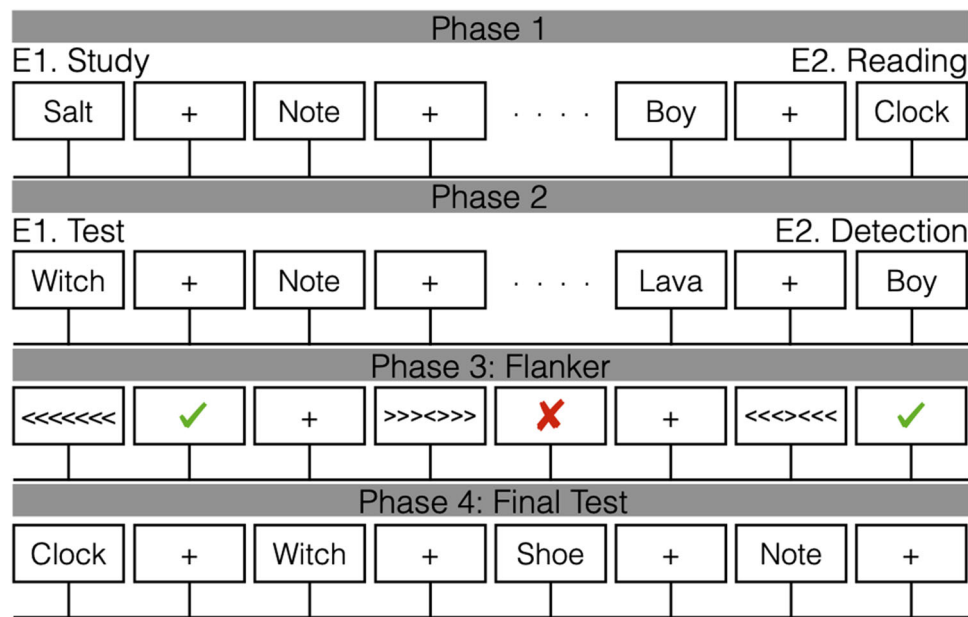


Figure 1. Task design. The phase 3 flanker task was divided into a practice subset of three runs and a main subset of six runs. The practice subset was completed prior to phase 1 to determine response duration during the main subset (see Methods). In E1 phase 1, participants studied individual words in anticipation of a later memory test. In E2 phase 1, participants read the words silently. In E1 phase 2, participants completed a recognition test and made old or new judgements using a confidence rating scale from 1 to 4, with 1 being definitely new and 4 being definitely old. In E2 phase 2, participants completed a detection phase in which the goal was to detect new words that were not presented in phase 1. Participants made old or new judgements without the use of a confidence rating scale. All participants then completed phase 3, a flanker task, in which they made speeded responses to a central target. Immediately after each response, a green check mark indicating a correct response or a red X indicating an incorrect response was presented. In phase 4, participants completed a final recognition memory test in which all the words from phases 1 and 2 were presented along with novel lures. Participants made old or new judgements using a confidence rating scale from 1 to 4, with 1 being definitely new and 4 being definitely old.

Phase 3. Prior to beginning phase 1, participants completed three practice runs of a flanker task in which they made speeded responses to a central target in a string of congruent (e.g., >>>>>>) or incongruent (e.g., <<<<<<) arrows. Feedback was presented immediately after each response as either a green check mark indicating a correct response or a red X indicating an incorrect response. Response duration, the interval in which a response was accepted, was initially set to 375 ms based on pilot data. To maintain difficulty and ensure an approximately balanced number of correct and incorrect responses during the main subset of phase 3, response duration was individually adjusted based on participants' accuracy following each practice run. If accuracy was below 50%, response duration increased by 25 ms, if accuracy was above 50%, response duration decreased by 25 ms. Thus, after completing the three practice runs, the final response duration could be a minimum of 300 ms and a maximum of 450 ms. After completing phase 2, participants completed the main subset of phase 3 which consisted of six runs of the flanker task. Throughout the main subset of phase 3, the response duration was fixed to that obtained from the final practice run. As our analyses focus on phase 2, we do not consider the flanker data further.

Phase 4. Participants completed a final recognition memory test in which all the words from phases 1 and 2 were presented along with novel lures. Trials were self-paced and participants made old or new judgements for each word using a confidence rating scale from 1 to 4, with 1 being definitely new and 4 being definitely old. Trials were separated by a 1,000 ms ISI. There were a total of 640 test trials with all 320 phase 2 words presented along with 320 novel lures. As our analyses focus on phase 2, we do not consider the final test data further.

Independent recognition memory experiment

To control for differing demands with regard to confidence judgments, we also include data from a third recognition memory experiment ("E3") on which we have previously reported (Smith et al., 2024). All of the analyses and results described here are new. Briefly, in E3, participants studied twelve 16-item word lists in anticipation of a later memory test. During the test phase, participants made old/new recognition memory judgments to targets and lures. We have reproduced the E3 instructions below (omitting instructions regarding sitting still, taking breaks, for brevity).

Phase 1 equivalent. During this experiment, you will be studying lists of words. Your task is to try to remember the words for a memory test.... In each run, you will view individual words and have 2 seconds to study each word. There will be fixation crosses (plus signs) shown after each item. You won't make a response on the keyboard, we just want you to look at the word, think about it, and try to remember it.

Phase 2 equivalent. During the test phase, you will see all of the words you studied from both phases along with new words that you haven't seen before. On each trial you will be presented with a word. The word could be one you've seen before or it could be an entirely new word. If you think you've seen the word before, press the K/D key with your right index finger. If you think this is the first time you've seen the word, press the D/K key with your left index finger.

Mnemonic state task

An independent group of participants completed a mnemonic state task. Participants were biased via explicit instructions on a trial-by-trial basis to engage an encoding or retrieval state, while perceptual input and behavioral demands were held constant. In this mnemonic state task (for specific study parameters, please see references Smith et al., 2022; Hong et al., 2023), participants viewed two lists of object images. For the first list, each object was new. For the second list, each object was again new but was categorically related to an object from the first list. For example, if list 1 contained an image of a bench, list 2 would contain an image of a different bench. During list 1, participants were instructed to encode each new object. During list 2, however, each trial contained an instruction to either encode the current object (e.g., the new bench) or to retrieve the corresponding object from list 1 (the old bench). Each object was presented for 2,000 ms. Participants completed either a two-alternative forced choice recognition test or a recency test on the object stimuli. We used the stimulus-locked list 2 data to train a multivariate pattern classifier (see below) to distinguish encoding and retrieval states.

EEG data acquisition and preprocessing

EEG recordings were collected using a BrainVision system and an ActiCap equipped with 64 Ag/AgCl active electrodes positioned

according to the extended 10–20 system. All electrodes were digitized at a sampling rate of 1,000 Hz and were referenced to electrode FCz. Offline, electrodes were later converted to an average reference. Impedances of all electrodes were kept below 50 k Ω . Electrodes that demonstrated high impedance or poor contact with the scalp were excluded from the average reference; however, these electrodes were included in all subsequent analysis steps. Bad electrodes were determined by voltage thresholding (see below).

Custom python codes were used to process the EEG data. We applied a high pass filter at 0.1 Hz, followed by a notch filter at 60 Hz and harmonics of 60 Hz to each participant's raw EEG data. We then performed three preprocessing steps (Nolan et al., 2010) to identify electrodes with severe artifacts. First, we calculated the mean correlation between each electrode and all other electrodes as electrodes should be moderately correlated with other electrodes due to volume conduction. We z-scored these means across electrodes and rejected electrodes with z-scores less than -3 . Second, we calculated the variance for each electrode, as electrodes with very high or low variance across a session are likely dominated by noise or have poor contact with the scalp. We then z-scored variance across electrodes and rejected electrodes with a $|z| \geq 3$. Finally, we expect many electrical signals to be autocorrelated, but signals generated by the brain versus noise are likely to have different forms of autocorrelation. Therefore, we calculated the Hurst exponent, a measure of long-range autocorrelation, for each electrode and rejected electrodes with a $|z| \geq 3$. Electrodes marked as bad by this procedure were excluded from the average re-reference. We then calculated the average voltage across all remaining electrodes at each time sample and re-referenced the data by subtracting the average voltage from the filtered EEG data. We used wavelet-enhanced independent component analysis (Castellanos and Makarov, 2006) to remove artifacts from eyeblinks and saccades.

EEG data analysis

We applied the Morlet wavelet transform (wave number 6) to the entire EEG time series across electrodes, for each of 46 logarithmically spaced frequencies (2–100 Hz; Long and Kahana, 2015), across all experiments. After log-transforming the power, we downsampled the data by taking a moving average across 100 ms time intervals from $-1,000$ to $3,000$ ms relative to the response and sliding the window every 25 ms, resulting in 157 time intervals (40 non-overlapping). Mean and standard deviation power were calculated across all trials and across time points for each frequency. Power values were then z-transformed by subtracting the mean and dividing by the standard deviation power. We followed the same procedure for the mnemonic state task, with 317 overlapping (80 non-overlapping) time windows from $4,000$ ms preceding to $4,000$ ms following stimulus onset (Smith et al., 2022).

Pattern classification analyses

Pattern classification analyses were performed using penalized (L2) logistic regression implemented via the sklearn module (0.24.2) in Python and custom Python code. For all classification analyses, classifier features were composed of spectral power across 63 electrodes and 46 frequencies. Before pattern classification analyses were performed, an additional round of z-scoring was performed across features (electrodes and frequencies) to eliminate trial-level differences in spectral power (Kuhl and Chun, 2014; Long and Kuhl, 2018; Smith et al., 2022). Therefore, mean univariate activity was matched precisely across all conditions and trial types. We extract “classifier evidence”, a continuous value reflecting the logit-transformed probability that the classifier assigns the correct mnemonic label (encode and retrieve) for each trial. Classifier evidence is used as a trial-specific, continuous measure of memory state information, which is used to assess the degree of retrieval state evidence present during hit and correct rejection (CR) trials.

Cross-study memory state classification

To measure retrieval state engagement in E1, E2, and E3, we conducted three stages of classification using the same methods as in our prior work (Long, 2023). First, we conducted within participant leave-one-run-out cross-validated classification (penalty parameter = 1) on all participants who completed the mnemonic state task ($N=103$; see Hong et al., 2023 for details). Four additional participants were included who did

not have resting state data (precluding inclusion in the cited study), but did have mnemonic state data. The classifier was trained to distinguish encoding versus retrieval states based on spectral power averaged across the 2,000 ms stimulus interval during list 2 trials. For each participant, we generated true and null classification accuracy values. We permuted condition labels (encode and retrieve) for 1,000 iterations to generate a null distribution for each participant. Any participant whose true classification accuracy fell above the 90th percentile of their respective null distribution was selected for further analysis ($N=37$; 28 female; age range = 18–42, mean age = 21.24 years). This thresholding reduces the degree to which noise contributes to the training data by excluding participants who do not have robust spectral dissociations between encode and retrieve trials. Second, we conducted leave-one-participant-out cross-validated classification (penalty parameter = 0.0001) on the selected participants to validate the mnemonic state classifier and obtained classification accuracy of 60% which is significantly above chance ($t_{36} = 6.000$, $p < 0.0001$), indicating that the cross-subject mnemonic state classifier is able to distinguish encoding and retrieval states. Finally, we applied the cross-subject mnemonic state classifier to the phase 2 trials of E1 and E2, and to the test phase trials of E3, specifically in 100 ms intervals from 500 ms pre-response to 500 ms post-response. We extracted classifier evidence, the logit-transformed probability that the classifier assigned a given phase 2 trial a label of encoding or retrieval. This approach provides a trial-level estimate of retrieval state evidence during hit and CR trials.

Statistical analyses

We used an independent samples *t*-test to assess the difference in CR rate between E1 and E2. We used mixed effects ANOVAs and *t*-tests to assess the effect of experiment (E1 and E2), response (hit and CR), and time interval on memory state evidence. We used a mixed effects ANOVA and *t*-tests to assess the effect of experiment (E2 and E3) and response (hit and CR) on memory state evidence during the pre-response interval. We used a repeated measures ANOVA and *t*-tests to assess the effect of response confidence (high and low) and time interval on memory state evidence for hits in E1. We used false discovery rate (FDR) to correct for multiple comparisons for post hoc *t*-tests (Benjamini and Hochberg, 1995).

Code accessibility

All raw, de-identified data and the associated experimental and analysis codes used in this study are available via the Long Term Memory Lab Website (longtermmemorylab.com).

Results

Top-down goals modulate test phase reaction times

Our general expectation is that the task instructions to either recognize old items (E1) or detect new items (E2) will impact the processes in which participants engage when responding to the test stimuli. The implicit assumption is that “old” is equal to “not-new”, but behavioral evidence (Brainerd et al., 2021) suggests that recognizing old items is not equivalent to detecting new items. To the extent that encoding specificity—the match between test goals and test probes—drives memory judgments, we would expect to find greater accuracy when the goal and probe type match (e.g., goal is to detect new and probe type is lure) compared to when the goal and probe type do not match (e.g., goal is to detect new and probe type is target). Alternatively, to the extent that participants use general or gist level information to reject items, we would expect to find greater accuracy when the goal and probe type do not match compared to when the goal and probe type match.

Based on our pre-registered pilot data, we expected participants in E2 to be more conservative, whereby their CR rates would be lower than those in E1. However, we did not find a significant difference in CR rates ($t_{70} = 0.3977$, $p = 0.6921$, $d = 0.0951$) between E1 ($M = 0.8141$, $SD = 0.1499$) and E2 ($M = 0.8273$, $SD = 0.1268$; Fig. 2A).

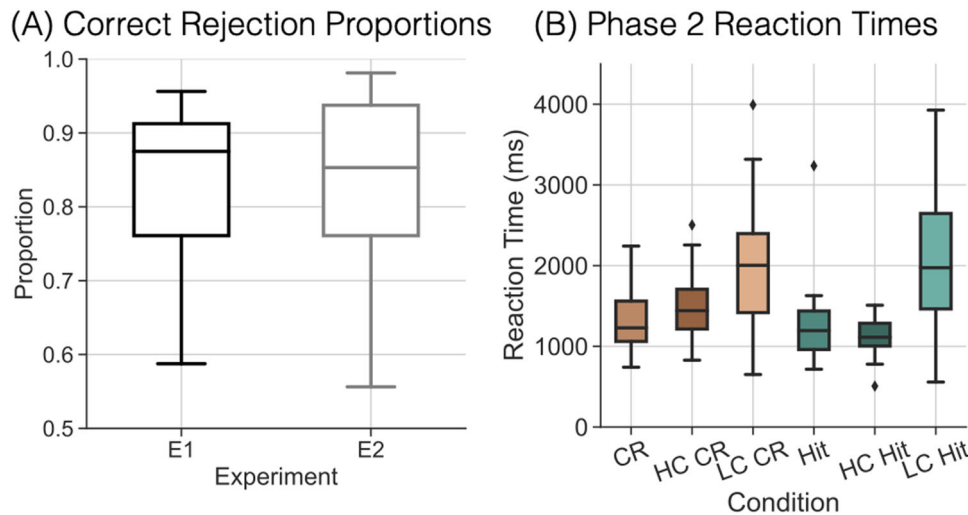


Figure 2. Test phase instructions modulate reaction times (RTs). **A**, Correct rejection (CR) proportions for E1 (black) and E2 (gray). We do not find a significant difference in CR rates across experiments. **B**, RTs for E2 CRs (orange) and hits (teal). For E1, we divided CRs and hits into high confident (HC; dark orange/teal) and low confident (LC; light orange/teal) responses. RTs are significantly faster for E2 CRs compared to E1 HC CRs and numerically faster for E1 HC hits compared to E2 hits. Box-and-whisker plots show median (center line), upper and lower quartiles (box limits), $1.5\times$ interquartile range (whiskers) and outliers (diamonds).

As the lack of difference in CR rates is difficult to interpret, we conducted an exploratory analysis of reaction times (RTs). Because participants in E1 made confidence judgments whereas participants in E2 did not, we divided response types into six conditions: CRs (E2), high confidence CRs (E1, response of “1”) and low confidence CRs (E1, responses of “2”), hits (E2), high confidence hits (E1, response of “4”) and low confidence hits (E1, responses of “3”; Fig. 2B). For a direct comparison between experiments, we excluded low confidence responses and conducted a 2×2 mixed effects ANOVA with factors of experiment (E1 and E2) and response type (hit and CR). We do not find a significant main effect of experiment ($F_{1,70} = 0.137$, $p = 0.712$, $\eta_p^2 = 0.002$). We find a significant main effect of response type ($F_{1,70} = 42.91$, $p < 0.0001$, $\eta_p^2 = 0.38$) and a significant interaction between response type and experiment ($F_{1,70} = 22.05$, $p < 0.0001$, $\eta_p^2 = 0.24$). This interaction was driven by significantly faster CRs in E2 ($M = 1302$ ms, $SD = 328.1$ ms) than high confidence CRs in E1 ($M = 1479$ ms, $SD = 379.6$ ms; $t_{70} = 2.080$, $p = 0.0411$, $d = 0.4973$) and numerically faster high confidence hits in E1 ($M = 1123$ ms, $SD = 200.3$ ms) than hits in E2 ($M = 1,244$ ms, $SD = 421.1$ ms; $t_{70} = 1.529$, $p = 0.1308$, $d = 0.3655$). These findings are consistent with the encoding specificity account. When the participant’s goal (e.g., detect new) matches the probe type (e.g., lure), responses are facilitated. These results demonstrate that the manipulated task goals impact recognition memory responses.

Top-down goals modulate retrieval state engagement

Our central goal was to test the hypothesis that top-down and bottom-up factors do not interact and specifically that top-down demands fully engage the retrieval state. We consider stimulus repetition to be a bottom-up factor, here represented by the memory response whereby hits constitute a repetition and CRs do not. We expect hits to induce a retrieval state in both E1 and E2. We consider test phase instructions to be a top-down factor. We expect that the demand to recognize old items (E1) will induce a retrieval state, *regardless of the actual probe*, whereas the demand to detect new items (E2) will not induce a retrieval state. To the extent that these top-down and bottom-up factors do not interact, we expect to find differential levels of retrieval state evidence specifically for CRs between E1 and E2 with no difference

in retrieval state evidence for hits. Alternatively, if top-down and bottom-up factors interact, we expect to find greater retrieval state evidence for hits in E1, whereby the combination of explicit demand to retrieve coupled with automatic repetition-driven retrieval will produce greatest engagement of the retrieval state.

To test our hypothesis, we applied a cross-study classifier to the response-locked whole-brain spectral data in each experiment and extracted retrieval state evidence from 500 ms preceding to 500 ms following the response (Fig. 3). Following our pre-registration, we conducted a $2\times 2\times 10$ mixed effects ANOVA with factors of response (hit and CR), experiment (E1 and E2), and time interval. We do not find a significant main effect of experiment ($F_{1,70} = 0.0019$, $p = 0.9657$, $\eta_p^2 < 0.0001$) or response type ($F_{1,70} = 0.339$, $p = 0.5623$, $\eta_p^2 = 0.0048$). We find a significant main effect of time interval ($F_{9,630} = 5.444$, $p < 0.0001$, $\eta_p^2 = 0.0722$), with generally more retrieval state evidence around the time of the response. We do not find a significant interaction between response type and time ($F_{9,630} = 0.848$, $p = 0.572$, $\eta_p^2 = 0.012$), response type and experiment ($F_{1,70} = 1.777$, $p = 0.1869$, $\eta_p^2 = 0.0248$), or time and experiment ($F_{9,630} = 1.579$, $p = 0.1179$, $\eta_p^2 = 0.0221$). We find a significant three-way interaction between response type, time, and experiment ($F_{9,630} = 2.174$, $p = 0.0222$, $\eta_p^2 = 0.0301$). We performed post hoc two sample t -tests comparing E1 and E2 at each time point separately for CRs and hits. We find significantly greater retrieval state evidence during E1 compared to E2 CRs in the -500 to -400 ms and -400 to -300 ms intervals (-500 to -400 : $t_{70} = 3.634$, $p = 0.0005$, $d = 0.8687$; -400 to -300 : $t_{70} = 3.234$, $p = 0.0019$, $d = 0.7731$; FDR corrected). All other post hoc comparisons did not survive multiple comparisons correction ($t_s < 1.914$, $p_s > 0.0597$).

We had expected that top-down demands could dissociate the two experiments early relative to the response onset and our initial finding of significant dissociations in retrieval state evidence in the 500 – 300 ms preceding the response is consistent with this expectation. Our next step was to directly test the specificity of the timing of this effect. Following our pre-registration, we averaged retrieval state evidence across the pre-response (-500 to 0 ms) and post-response (0 – 500 ms) intervals. We conducted a 2×2 mixed effects ANOVA with factors of

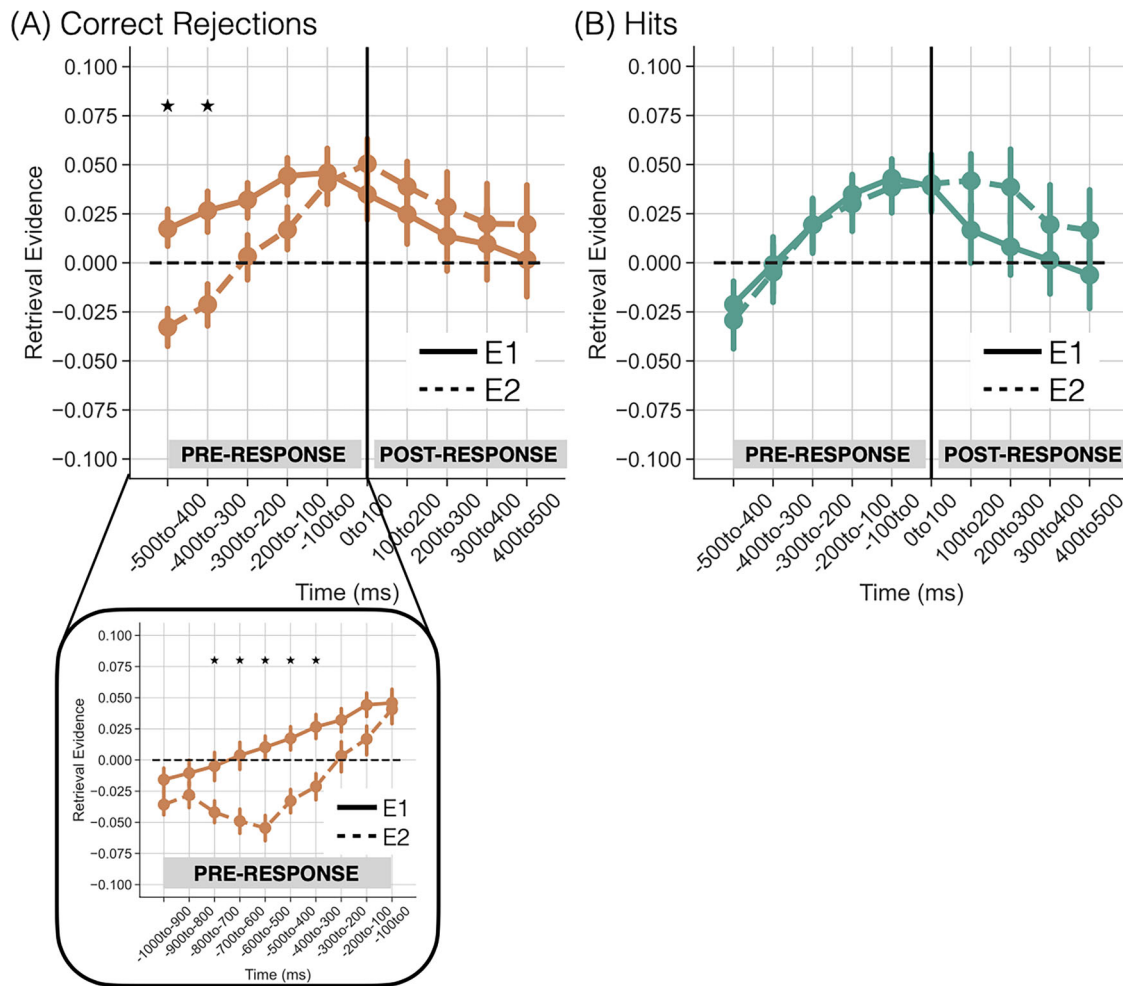


Figure 3. Top-down demands modulate retrieval state engagement. We applied a cross-study classifier to response-locked test phase data to measure retrieval state evidence separately for CRs and hits in E1 (solid) and E2 (dashed). Positive y-axis values indicate greater retrieval state evidence. The vertical line at time 0 to 100 ms indicates the onset of the response. **A**, Retrieval state evidence preceding the response is significantly greater for CRs in E1 compared to E2. **Lower panel.** When we extend the pre-response interval, we find greater retrieval state evidence for E1 compared to E2 in the 800–300 ms preceding the response. **B**, There is no significant difference in retrieval state evidence for hits across experiments. Stars reflect paired *t*-tests that survived false discovery rate (FDR) correction. Error bars reflect standard error of the mean.

interval (pre-response and post-response) and experiment (E1 and E2). We do not find a significant main effect of experiment ($F_{1,70} = 0.4657$, $p = 0.4972$, $\eta_p^2 = 0.0066$) or interval ($F_{1,70} = 0.3475$, $p = 0.5574$, $\eta_p^2 = 0.0049$). We find a marginal interaction between interval and experiment ($F_{1,70} = 3.974$, $p = 0.0501$, $\eta_p^2 = 0.0537$), driven by a greater dissociation between experiments in the pre-response compared to the post-response interval. This dissociation in time between experiments is consistent with the expectation that top-down demands should engage the retrieval state early relative to response onset.

As it is possible that the observed dissociations in CRs extend even earlier, prior to the *a priori* selected pre-response window of –500 to 0 ms, we conducted an exploratory analysis in which we extended the pre-response interval and measured retrieval state evidence during E1 and E2 CRs for the 1,000 ms preceding the response (Fig. 3A, lower panel). We conducted a 2×10 mixed effects ANOVA with factors of experiment (E1 and E2) and time interval. We find a significant main effect of experiment ($F_{1,70} = 11.61$, $p = 0.0011$, $\eta_p^2 = 0.1423$), driven by greater retrieval state evidence for E1 compared to E2. We find a significant main effect of time interval ($F_{9,630} = 21.63$, $p < 0.0001$, $\eta_p^2 = 0.2361$), driven by greater retrieval state evidence immediately preceding the response. We find a significant interaction

between experiment and time interval ($F_{9,630} = 2.955$, $p = 0.0019$, $\eta_p^2 = 0.0405$), driven by greater retrieval state evidence for E1 compared to E2 for the intervals spanning 800–300 ms preceding the response (Table 1; FDR corrected). These results show that the retrieval state is engaged several hundred milliseconds prior to a CR specifically when participants' goal is to recognize old items.

Considering the results for both hits and CRs together, we show that top-down task demands can lead to the recruitment of the retrieval state independent from bottom-up stimulus repetition. That we do not find an additive or over-additive interaction between experiment and response suggests that the combination of the demand to retrieve and stimulus repetition does not further enhance retrieval state engagement beyond either factor alone. Instead, these results suggest that the retrieval state may be engaged any time participants are explicitly directed to retrieve a past experience, regardless of the actual status of the test probe.

Top-down demands induce the retrieval state in the absence of confidence judgments

Given our motivation to test the hypothesis that a retrieval state engaged by top-down versus bottom-up factors should differentially relate to the content that is or is not retrieved, participants

Table 1. Post hoc *t*-tests comparing pre-response retrieval state evidence during CRs in E1 and E2

Time interval (ms)	E1 versus E2		
	t_{70}	p	d
–1000 to –900	1.488	0.1412	0.3558
–900 to –800	1.218	0.2275	0.2911
–800 to –700	2.569	0.0123	0.6141
–700 to –600	3.686	0.0004	0.8811
–600 to –500	4.933	<0.0001	1.179
–500 to –400	3.634	0.0005	0.8687
–400 to –300	3.234	0.0019	0.7731
–300 to –200	1.914	0.0597	0.4575
–200 to –100	1.838	0.0704	0.4393
–100 to 0	0.3005	0.7647	0.0718

Note: Bold values indicate tests that survive FDR correction.

were instructed to provide a confidence judgment specifically for E1. This manipulation further differentiates E1 from E2 and may thus underlie the observed retrieval state dissociations reported above. Namely, greater retrieval evidence for CRs in E1 may reflect the demand to make a confidence judgment, rather than the demand to recognize old items.

To rule out confidence as an alternative explanation for our findings, we conducted an exploratory analysis in which we re-analyzed an independently collected dataset from our lab (Smith et al., 2024; “E3”). In E3, participants performed a traditional old/new recognition memory task on lists of words. The instructions in E3 were identical to those in E1 phases 1 and 2 (Fig. 1) except that, critically, there was no demand to provide a confidence judgment. Thus the response options (“old” and “new”) were matched across E2 and E3 and the only difference was the task goal to either detect new items (E2) or recognize old items (E3). To the extent that the demand to recognize old items induces retrieval, we should observe the same pattern when comparing E2 to E3 as when E2 is compared to E1—namely, elevated retrieval state evidence during E3 compared to E2 CRs. We applied our cross-study classifier to the response-locked test phase data in E3.

To the extent that the top-down demand to recognize old items induces a retrieval state, we should again observe an experiment by response interaction, whereby pre-response retrieval evidence is greater for CRs in E3 relative to E2. In our pre-registration, we planned to separately consider pre- and post-response intervals in the event that we found interactions with time interval, thus we averaged retrieval state evidence across the 500 ms pre-response interval (Fig. 4). We conducted a 2×2 mixed effects ANOVA with factors of response type (hit and CR) and experiment (E2 and E3). We do not find a significant main effect of experiment ($F_{1,72} = 0.6562$, $p = 0.4206$, $\eta_p^2 = 0.009$) or response type ($F_{1,72} = 0.8406$, $p = 0.3623$, $\eta_p^2 = 0.0115$). We find a significant interaction between response type and experiment ($F_{1,72} = 6.032$, $p = 0.0165$, $\eta_p^2 = 0.0773$). This interaction was driven by numerically greater retrieval evidence for CRs in E3 ($M = 0.257$, $SD = 0.0521$) compared to E2 ($M = 0.0014$, $SD = 0.0567$; $t_{72} = 1.896$, $p = 0.0619$, $d = 0.4466$). The difference in retrieval evidence between hits in E3 ($M = 0.0065$, $SD = 0.0475$) and E2 ($M = 0.0107$, $SD = 0.0733$) was not significant ($t_{72} = 0.2945$, $p = 0.7692$, $d = 0.0691$). The dissociation between E2 and E3 mirrors that of E2 and E1, whereby we find a significant response by experiment interaction ($F_{1,70} = 4.255$, $p = 0.048$, $\eta_p^2 = 0.0573$), driven by significantly greater retrieval evidence for CRs in E1 ($M = 0.033$, $SD = 0.0519$) compared to E2 ($t_{72} = 2.453$, $p = 0.0167$, $d = 0.5864$).

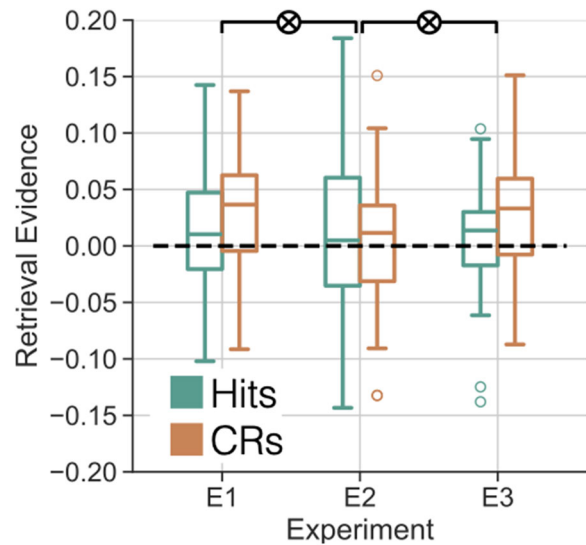


Figure 4. Validation of the modulation of top-down demands on retrieval state engagement. We applied a cross-study classifier to response-locked test phase data in three separate experiments. We extracted average retrieval state evidence in the 500–0 ms preceding hits (teal) and CRs (orange). In E1, participants are instructed to recognize old items while making confidence judgments. In E2, participants are instructed to detect new items. In E3, participants are instructed to recognize old items without making confidence judgments. We find a significant response by experiment interaction between E1 and E2 ($p = 0.0428$) and between E2 and E3 ($p = 0.0165$). Box-and-whisker plots show median (center line), upper and lower quartiles (box limits), 1.5x interquartile range (whiskers), and outliers (circles).

In sum, we validated that the dissociation of retrieval evidence during CRs across experiments is not due to the demand to make a confidence judgment. We find that when participants are instructed to recognize old items—classic recognition memory task instructions—the retrieval state is recruited during CRs, whereas when participants are instructed to detect new items, no such engagement is observed. This result suggests that the top-down goal to retrieve induces a retrieval state even in instances when the probe itself (a lure) may not automatically induce retrieval.

Top-down retrieval state engagement is modulated by lack of retrieved content

Given that top-down demands engage the retrieval state in the absence of bottom-up factors, our second major goal was to test the extent to which the retrieval state is engaged by the presence or absence of retrieved content. Our initial investigation across experiments E1, E2, and E3 suggests that the retrieval state may be driven by the absence of retrieved content; however, analysis of CRs is ambiguous. Theoretically, there is no retrievable content for CRs, given that the probe is a lure. However, participants may evaluate evidence to make a decision for lures. Hits provide a clearer assessment of retrieved content as the assumption is that participants reinstate study-phase content during the test phase in order to make a decision for targets (Danker and Anderson, 2010). The content-present versus content-absent accounts make opposite predictions for retrieval state engagement during hits made with high versus low confidence. High confidence responses are thought to be supported by greater reinstatement or overall evidence, relative to low confidence responses (Baldson et al., 2020). Therefore, to the extent that the retrieval state tracks retrieved content, there should be more retrieval state evidence during high compared to low confidence hits. Alternatively, to the extent that the retrieval state

tracks attention or control required to access a representation, there should be more retrieval state evidence during low compared to high confidence hits. The logic is that less evidence should require more internal attention to support a decision, possibly by prolonging the search or evidence accumulation process (Callaway et al., 2023).

We tested the content-present versus content-absent accounts by assessing retrieval state evidence as a function of confidence specifically for E1 hits (Fig. 5A). Following our pre-registration, we conducted a 2×10 repeated measures ANOVA with factors of response confidence (high and low) and time interval (−500 to 500 in 100 ms intervals, response-locked). We find a significant main effect of response confidence ($F_{1,35} = 7.085$, $p = 0.0117$, $\eta_p^2 = 0.1684$) driven by greater retrieval state evidence for low ($M = 0.0432$, $SD = 0.0798$) compared to high ($M = 0.0078$, $SD = 0.0526$) confidence hits. We find a significant main effect of time interval ($F_{9,315} = 3.722$, $p = 0.0002$, $\eta_p^2 = 0.961$), driven by greater pre-response compared to post-response retrieval evidence. We find a significant interaction between response confidence and time interval ($F_{9,315} = 7.618$, $p < 0.0001$, $\eta_p^2 = 0.1787$), whereby retrieval state evidence is significantly greater pre-response for low ($M = 0.0729$, $SD = 0.0769$) compared to high ($M = 0.0045$, $SD = 0.0744$) confidence hits ($t_{35} = 4.104$, $p = 0.0002$, $d = 0.905$). Post-response retrieval state evidence does not significantly differ between low ($M = 0.0135$, $SD = 0.12$) and high ($M = 0.0111$, $SD = 0.0742$) confidence hits ($t_{35} = 0.1643$, $p = 0.8705$, $d = 0.0241$). We performed post hoc paired sample t -tests comparing high and low confidence hits at each time point. We find significantly greater retrieval state evidence during low compared to high confidence hits in the first three pre-response intervals (−500 to −400: $t_{35} = 5.145$, $p < 0.0001$, $d = 1.023$; −400 to −300: $t_{35} = 3.999$, $p = 0.0003$, $d = 0.8972$; −300 to −200: $t_{35} = 3.701$, $p = 0.0007$, $d = 0.7862$; FDR corrected). All other post hoc comparisons did not survive multiple comparisons correction (t 's < 2.434 , p 's > 0.0202). Together, these results provide support for the content-absent account whereby the retrieval state is more strongly engaged during trials in which less evidence is available to make a decision.

Elevated retrieval state evidence for low compared to high confidence hits may reflect a general effect of difficulty, rather than a memory-specific response. That is, there may be increased retrieval state evidence preceding difficult decisions regardless of top-down goals. To test this alternative, we conducted an exploratory analysis of retrieval state evidence as a function of slow versus fast hits in E2 and E3. The logic is that slow hits should show more retrieval state evidence than fast hits in both experiments if retrieval state evidence tracks the difficulty of responding (whereby slower responses correspond to more difficult decisions). However, to the extent that the retrieval state reflects the attempt to retrieve, slow and fast hits should only dissociate in E3, as there should be no attempt to retrieve in E2 given the task instructions.

We separately assessed retrieval state evidence for slow and fast hits in E2 and E3 (Fig. 5B,C). To divide the trials based on RTs, we calculated the median RT across all correct responses (hits and CRs) for each participant, and labeled hits as either “fast” (those below the median RT) or “slow” (those above the median RT; Smith et al., 2024). We conducted two 2×10 repeated measures ANOVA with factors of response (slow and fast) and time interval (−500 to 500 in 100 ms intervals, response-locked). In E2, we do not find a significant main effect of response ($F_{1,35} = 3.588$, $p = 0.0665$, $\eta_p^2 = 0.093$). We find a significant main effect of time ($F_{9,315} = 2.936$, $p = 0.0023$, $\eta_p^2 = 0.0774$). We do not find a significant interaction between response and time ($F_{9,315} = 0.3614$, $p = 0.9525$, $\eta_p^2 = 0.0102$). Bayes factor analysis revealed that a model without response (H_0) is preferred to a model with response (H_1 ; $BF_{10} = 0.0014$, extreme evidence for H_0). In E3, we find a significant main effect of response ($F_{1,37} = 63.24$, $p < 0.0001$, $\eta_p^2 = 0.6309$), driven by greater retrieval evidence for slow ($M = 0.0697$, $SD = 0.0491$) compared to fast ($M = -0.0115$, $SD = 0.0471$) hits. We find a significant main effect of time ($F_{9,333} = 4.143$, $p < 0.0001$, $\eta_p^2 = 0.1007$). We find a significant interaction between response and time ($F_{9,333} = 11.20$, $p < 0.0001$, $\eta_p^2 = 0.2323$), driven by greater retrieval evidence for slow compared to fast hits in the pre-response interval ($M = 0.1056$, $SD = 0.079$) compared to the post-response interval

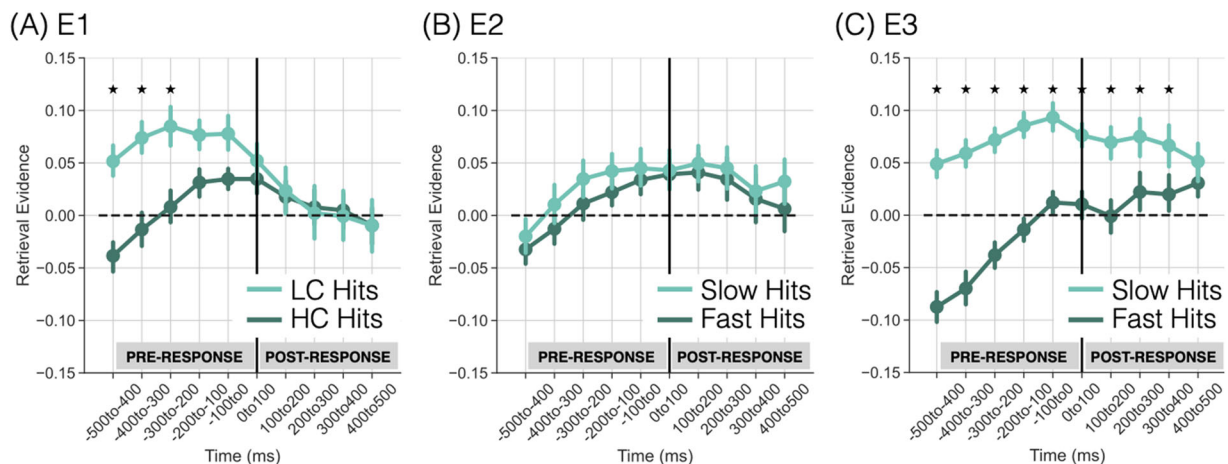


Figure 5. Confidence during successful retrieval modulates retrieval state engagement. We applied a cross-study classifier to response-locked test phase data in E1, E2, and E3 across ten time intervals from 500 ms preceding to 500 ms following the response. We measured retrieval state evidence for high confidence hits (dark teal) and low confidence hits (light teal) in E1 (A) and for fast hits (dark teal) and slow hits (light teal) in E2 (B) and E3 (C). **A**, We find significantly greater retrieval state evidence for low compared to high confidence hits in the 500–200 ms intervals preceding the response (p 's < 0.0007). **B**, We do not find a significant difference in retrieval state evidence for slow versus fast hits in E3 ($p = 0.0665$). **C**, We find greater retrieval state evidence for slow compared to fast hits in all time intervals except the final 400–500 ms interval (p 's < 0.0004). Stars reflect paired t -tests that survived FDR correction. Error bars reflect standard error of the mean.

($M = 0.0512$, $SD = 0.0589$). We performed post hoc paired sample t -tests comparing fast and slow hits at each time point. We find significantly greater retrieval state evidence during slow compared to fast hits in all time intervals (t 's > 3.870 , p 's < 0.0004 ; FDR corrected) except the 400–500 ms interval ($t_{35} = 1.302$, $p = 0.201$, $d = 0.1935$; FDR corrected).

Together, these results provide support for the content-absent account whereby the retrieval state is more strongly engaged during trials in which less evidence is available to make a recognition decision.

Discussion

The aim of this study was to identify the joint impact of bottom-up and top-down factors on retrieval state engagement. We conducted two independent recognition memory experiments in which we manipulated top-down goals to either recognize old items (E1) or detect new items (E2). We included a third experiment (E3) to control for possible effects of making a confidence judgment. We recorded scalp EEG and used a cross-study decoding approach (Long, 2023) to measure response-locked retrieval state engagement during the test phase of each experiment. Retrieval state engagement was greater during CRs when participants' goal was to recognize old items (E1 and E3) compared to when participants' goal was to detect new items (E2). Furthermore, retrieval state engagement was greater for low, compared to high confidence hits in E1, and for slow compared to fast hits in E3, when retrieved content is presumably low. Together, these findings suggest that the explicit top-down demand to remember the past recruits the retrieval state independent from bottom-up factors such as stimulus repetition. This intentional retrieval state engagement may reflect internally directed attention deployed to accomplish the task goal of recognizing old items.

We manipulated participants' task goals to either identify old items (E1) or identify new items (E2). Prior behavioral work has shown that these two goals produce different behavioral outcomes, despite their surface-level similarity (Brainerd et al., 2021). Here we find that RTs are facilitated when the test phase goals match the probe—RTs are faster for hits in E1 and for CRs in E2. These effects can be explained by the encoding specificity principle (Thomson and Tulving, 1970) whereby a match between the goal (e.g., recognize old) and the test probe (e.g., target) enables enhanced responding. Interestingly, the prior work found the opposite pattern whereby memory accuracy was better when the goal (e.g., recognize old) did not match the test probe (e.g., lure). This difference may potentially be due to the relatively higher CR rates in the current study compared to the prior study. Ultimately, however, these findings indicate that the modest manipulation of asking participants to either recognize old items or detect new items is sufficient to shift participants' test phase goals.

We find evidence for a top-down-driven retrieval state that is independent from bottom-up factors. Whereas hits recruit the retrieval state regardless of task goals, CRs recruit the retrieval state when the task goal is to recognize old items. These findings are consistent with the conceptualization of the retrieval mode as an intentional state that is a precursor to, but separate from, retrieval success (Tulving, 1983). The notion is that one must enter a retrieval state in order to interpret a stimulus as a retrieval cue (Rugg and Wilding, 2000). Thus, in the present study, because participants were explicitly instructed to recognize old items in E1 and E3, they engaged a retrieval state and treated every test probe—including lures—as a retrieval cue.

Our interpretation that retrieval state engagement is driven by the demand or attempt to retrieve—and not by retrieval success—is supported by our finding that retrieval state evidence is greater for both low confidence hits in E1 and slow hits in E3. We consider slow versus fast hits in E3 as analogous to low versus high confidence hits in E1, insofar as high confidence responses are committed more quickly than low confidence responses (e.g., Weidemann and Kahana, 2016). If the retrieval state specifically reflected retrieval success, we would not expect to find a confidence-based modulation, as both trials constitute retrieval success. If anything, we might expect greater retrieval state evidence for high compared to low confidence hits, whereby stronger representations produce greater retrieval state engagement. Instead, we find greater retrieval state evidence for low compared to high confidence hits which may reflect differences in memory search or evidence accumulation across the two conditions (Baldson et al., 2020; Callaway et al., 2023; Lee et al., 2023). That we do not find a slow versus fast hit dissociation in E2 argues against the interpretation that the retrieval state is generally driven by task or response difficulty. Instead, our results are consistent with the interpretation that the retrieval state is engaged via top-down demands in service of the attempt to access an internally stored representation.

Our findings suggest that the retrieval state as measured here reflects internal attention engaged in the service of the attempt to retrieve. Both theoretical proposals and computational models suggest that long-term memory retrieval reflects a form of internally directed attention (Chun et al., 2011; Long et al., 2018; Logan et al., 2021). Internal attention and memory retrieval share a number of neural substrates, including recruitment of the default mode network (Smallwood et al., 2013; Kim, 2015; Buckner and DiNicola, 2019) and spectral signals such as theta and alpha power (Kam et al., 2019; Woodman et al., 2022). Our interpretation is that the instruction to recognize old items induces participants to turn their mind's eye inward when evaluating test probes. When more evidence is required to evaluate a probe, more internal attention should be recruited, which is consistent with our finding of greater retrieval state engagement during low compared to high confidence hits. We anticipate that under any circumstances in which participants are given and follow traditional memory task instructions to recognize or remember old items, the retrieval state will be recruited, regardless of the status of the test probe. There may be important individual differences in the ability to engage the retrieval state in the service of task goals, which will in turn impact downstream processing and behavior.

Our results suggest that the retrieval state is differentially engaged based on task goals. As part of our task manipulation, we also modified participants' task goals when initially encountering the stimuli. As E1 participants were explicitly instructed to intentionally encode the stimuli, this may have created a levels-of-processing difference (Craik and Lockhart, 1972; Craik and Tulving, 1975) such that stimuli in E1 might have been more deeply processed than those in E2. We would expect a levels-of-processing difference to specifically impact hits. Although we cannot draw strong conclusions from a null result, the lack of a difference in retrieval state evidence between E1 and E2 hits provides evidence against an account whereby study-phase levels-of-processing differences modulate test phase memory state engagement. However, we anticipate that different study-phase goals have the potential to modulate study-phase memory states and see this as an exciting avenue for future

work. Prior work has shown that shifts in top-down goals can change how stimuli are represented in the brain (Aly and Turk-Browne, 2016; Long and Kuhl, 2021; Lim et al., 2023). Our results extend these findings by showing that task goals modulate the processes or states that are engaged in the service of those task goals.

If the retrieval state is only engaged for CRs when participants recognize old items, what processes are engaged when participants are instructed to detect new items in E2? By one account, participants may engage in some form of novelty detection and greater univariate activity for lures versus targets may reflect a novelty signal (Knight, 1996; Tulving et al., 1996; Grunwald et al., 1998; Ranganath and Rainer, 2003; Duzel et al., 2010). Potentially, the detection of novelty may induce an encoding state (Patil and Duncan, 2018), governed by acetylcholine release. The design of our independent mnemonic state classifier is such that “negative” retrieval state evidence is equivalent to evidence for an encoding state. However, we view this as a strength, reflecting our assumption—based on rodent electrophysiological work (Hasselmo et al., 2002)—that encoding and retrieval exist along a continuum, due to their reliance on shared neural circuitry. We find evidence for encoding state engagement during CRs in E2 in the early time intervals preceding the response, which is consistent with a novelty/encoding interpretation. However, it is important to note that despite the differences leading into CRs, we do find evidence of retrieval state engagement immediately preceding all responses in both E1 and E2. We speculate that this retrieval state response may reflect the decision process, rather than a memory process per se. In prior work, we have found that retrieval state engagement “ramps up” leading to a behavioral decision in a spatial attention task (Long, 2023), suggesting that the retrieval state may be engaged to evaluate a decision in addition to supporting episodic memory retrieval attempts. Future work will be needed to specifically address how memory versus decision-making processes engage memory brain states.

Despite clear evidence that top-down demands engage the retrieval state, we still expect that bottom-up factors play a role in memory state engagement. We observe retrieval state engagement for hits in both E1 and E2, as would be expected from a bottom-up account; however, these effects are relatively modest. Our interpretation is that the retrieval state can be induced automatically via bottom-up inputs, but that these effects are likely to be largely masked by top-down factors and will only emerge either in the absence of top-down demands or when bottom-up factors are highly salient. Old versus new stimuli may lead the hippocampus into a pattern separating versus pattern completing state (Duncan et al., 2012; Patil and Duncan, 2018) and temporal contextual overlap between experiences may automatically induce the retrieval state (Smith et al., 2022). Our manipulation of bottom-up factors in the present study was minimal, with targets that were only repeated once, having been presented during study and during test. Thus, with more substantial manipulations—more repetitions, higher similarity to other stimuli, etc.—we might find greater influences of bottom-up factors on the retrieval state. An exciting future direction will be to investigate the full range of bottom-up factors that can induce retrieval state engagement, especially in the absence of top-down factors.

Taken together, our findings indicate that the retrieval state is engaged in the service of the top-down goal to recognize old items. The retrieval state may thus be recruited in response to any demand to direct attention internally, regardless of whether

stimuli themselves have previously been encountered. These results advance our understanding of the interplay between top-down and bottom-up factors on memory state engagement.

References

- Aly M, Turk-Browne NB (2016) Attention promotes episodic encoding by stabilizing hippocampal representations. *Proc Natl Acad Sci USA* 113:E420–E429.
- Balsdon T, Wyart V, Mamassian P (2020) Confidence controls perceptual evidence accumulation. *Nat Commun* 11:1753.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* 57:289–300.
- Brainerd CJ, Bialer DM, Chang M, Upadhyay P (2021) A fundamental asymmetry in human memory: old = not-new and new = not-old. *J Exp Psychol Learn Mem Cognit* 48:1850–1867.
- Buckner RL, DiNicola LM (2019) The brain’s default network: updated anatomy, physiology and evolving insights. *Nat Rev Neurosci* 20:593–608.
- Cabeza R, Mazuz YS, Stokes J, Kragel JE, Woldorff MG, Ciaramelli E, Moscovitch M (2011) Overlapping parietal activity in memory and perception: evidence for the attention to memory model. *J Cognit Neurosci* 23:3209–3217.
- Callaway F, Griffiths TL, Norman KA, Zhang Q (2023) Optimal metacognitive control of memory recall. *Psychol Rev* 131:781–811.
- Castellanos NP, Makarov VA (2006) Recovering EEG brain signals: artifact suppression with wavelet enhanced independent component analysis. *J Neurosci Methods* 158:300–312.
- Chun MM, Golomb JD, Turk-Browne NB (2011) A taxonomy of external and internal attention. *Ann Rev Psychol* 62:73–101.
- Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3:201–215.
- Craik FIM, Lockhart RS (1972) Levels of processing: a framework for memory research. *J Verbal Learn Verbal Behav* 11:671–684.
- Craik FIM, Tulving E (1975) Depth of processing and the retention of words in episodic memory. *J Exp Psychol Gen* 104:268–294.
- Danker JF, Anderson JR (2010) The ghosts of brain states past: remembering reactivates the brain regions engaged during encoding. *Psychol Bull* 136:87–102.
- Duncan KD, Sadanand A, Davachi L (2012) Memory’s penumbra: episodic memory decisions induce lingering mnemonic biases. *Science* 337:485–487.
- Düzel E, Penny WD, Burgess N (2010) Brain oscillations and memory. *Curr Opin Neurobiol* 20:143–149.
- Friendly M, Franklin PE, Hoffman D, Rubin DC (1982) The Toronto word pool: norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behav Res Methods Instrum* 14:375–399.
- Grunwald T, Lehnertz K, Heinze HJ, Helmstaedter C, Elger CE (1998) Verbal novelty detection within the human hippocampus proper. *PNAS* 95:3193–3197.
- Harris KD, Thiele A (2011) Cortical state and attention. *Nat Rev Neurosci* 12:509–523.
- Hasselmo ME, Bodelón C, Wyble BP (2002) A proposed function for hippocampal theta rhythm: separate phases of encoding and retrieval enhance reversal of prior learning. *Neural Comput* 14:793–817.
- Herron JE, Wilding EL (2004) An electrophysiological dissociation of retrieval mode and retrieval orientation. *NeuroImage* 22:1554–1562.
- Herron JE, Wilding EL (2006) Brain and behavioral indices of retrieval mode. *NeuroImage* 32:863–870.
- Hong Y, Smith DE, Moore IL, Long NM (2023) Spatiotemporal dynamics of memory encoding and memory retrieval states. *J Cognit Neurosci* 35:1463–1477.
- Kam JWY, Lin JJ, Solbakk AK, Endestad T, Larsson PG, Knight RT (2019) Default network and frontoparietal control network theta connectivity supports internal attention. *Nat Human Behav* 3:1263–1270.
- Kim H (2015) Encoding and retrieval along the long axis of the hippocampus and their relationships with dorsal attention and default mode networks: the HERNET model. *Hippocampus* 25:500–510.
- Knight RT (1996) Contribution of human hippocampal region to novelty detection. *Nature* 383:256–259.
- Kuhl BA, Chun MM (2014) Successful remembering elicits event-specific activity patterns in lateral parietal cortex. *J Neurosci* 34:8051–8060.

- Lee DG, Daunizeau J, Pezzulo G (2023) Evidence or confidence: what is really monitored during a decision? *Psychonomic Bull Rev* 30:1360–1379.
- Lepage M, Ghaffar O, Nyberg L, Tulving E (2000) Prefrontal cortex and episodic memory retrieval mode. *Proc Natl Acad Sci USA* 97:506–511.
- Lim YL, Lang DJ, Diana RA (2023) Cognitive tasks affect the relationship between representational pattern similarity and subsequent item memory in the hippocampus. *NeuroImage* 227:120241.
- Logan GD, Cox GE, Annis J, Lindsey DRB (2021) The episodic flanker effect: memory retrieval as attention turned inward. *Psychol Rev* 128:397–445.
- Long NM (2023) The intersection of the retrieval state and internal attention. *Nat Commun* 14:3861.
- Long NM, Kahana MJ (2015) Successful memory formation is driven by contextual encoding in the core memory network. *NeuroImage* 119:332–337.
- Long NM, Kuhl BA (2018) Bottom-up and top-down factors differentially influence stimulus representations across large-scale attentional networks. *J Neurosci* 38:2495–2504.
- Long NM, Kuhl BA (2019) Decoding the tradeoff between encoding and retrieval to predict memory for overlapping events. *NeuroImage* 201:116001.
- Long NM, Kuhl BA (2021) Cortical representations of visual stimuli shift locations with changes in memory states. *Curr Biol* 31:1119–1126.e5.
- Long NM, Kuhl BA, Chun MM (2018) Memory and attention. In: *Steven's handbook of experimental psychology and cognitive neuroscience* (Wixted JT, ed), Ed 4. John Wiley & Sons, Inc.
- Nobre AC (2018) Attention. In: *Steven's handbook of experimental psychology and cognitive neuroscience* (Wixted JT, ed), Ed 4. John Wiley & Sons, Inc.
- Nolan H, Whelan R, Reilly R (2010) FASTER: fully automated statistical thresholding for EEG artifact rejection. *J Neurosci Methods* 192:152–162.
- Patil A, Duncan K (2018) Lingering cognitive states shape fundamental mnemonic abilities. *Psychol Sci* 29:45–55.
- Polyn SM, Kahana MJ (2008) Memory search and the neural representation of context. *Trends Cognit Sci* 12:24–30.
- Ranganath C, Rainer G (2003) Neural mechanisms for detecting and remembering novel events. *Nat Rev Neurosci* 4:193–202.
- Ratcliff R, McKoon G (2008) The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput* 20:873–922.
- Rugg MD, Wilding EL (2000) Retrieval processing and episodic memory. *Trends Cognit Sci* 4:108–115.
- Smallwood J, Tipper C, Brown K, Baird B, Engen H, Michaels JR, Schooler JW (2013) Escaping the here and now: evidence for a role of the default mode network in perceptually decoupled thought. *NeuroImage* 69:120–125.
- Smith DE, Moore IL, Long NM (2022) Temporal context modulates encoding and retrieval of overlapping events. *J Neurosci* 42:3000–3010.
- Smith DE, Wheelock JR, Long NM (2024) Response-locked theta dissociations reveal potential feedback signal following successful retrieval. *Imaging Neurosci* 2:1–16.
- Tarder-Stoll H, Jayakumar M, Dimsdale-Zucker HR, Günseli E, Aly M (2020) Dynamic internal states shape memory retrieval. *Neuropsychologia* 138.
- Thomson DM, Tulving E (1970) Associative encoding and retrieval: weak and strong cues. *J Exp Psychol* 86:255–262.
- Tulving E (1983) *Elements of episodic memory*. New York: Oxford.
- Tulving E, Markowitsch HJ, Craik FI, Habib R, Houle S (1996) Novelty and familiarity activations in PET studies of memory encoding and retrieval. *Cereb Cortex* 6:71–79.
- Weidemann CT, Kahana MJ (2016) Assessing recognition memory using confidence ratings and response times. *R Soc Open Sci* 3:150670.
- Woodman GF, Wang S, Sutterer DW, Reinhart RMG, Fukuda K (2022) Alpha suppression indexes a spotlight of visual-spatial attention that can shine on both perceptual and memory representations. *Psychonomic Bull Rev* 29:681–698.