Neural Predictors of Subsequent Long-Term Memory After (De)prioritization in Working Memory

Frieda Born (born@mpib-berlin.mpg.de)[1,2,3,4]
Or Yizhar (yizhar@mpib-berlin.mpg.de)[1,4]
Bernhard Spitzer (spitzer@mpib-berlin.mpg.de)[1,4]

[1] Research Group Adaptive Memory and Decision Making, Max Planck Institute for Human Development, Berlin, Germany
[2] Machine Learning Group, Technical University Berlin, Germany
[3] BIFOLD, Berlin Institute for the Foundations of Learning and Data, Berlin, Germany
[4]Chair of Biopsychology, Technische Universität Dresden, Chemnitzer Straße 46, 01187, Dresden, Germany

Abstract

Which factors determine how well information held in working memory (WM) can later be remembered from long-term memory (LTM)? Prior behavioral work (presented at CCN 2024) from our lab suggests that active WM retrieval (WM-"testing"), particularly when retrieving information that has been deprioritized in WM, can enhance subsequent LTM performance; similar to the well-known "testing-effect" in LTM research. However, the neural mechanisms underlying these effects remain unknown. Here, we used fMRI to study which neural signatures of WM processing predict later LTM performance, so-called "subsequent memory effects" (SMEs). Using a dual-retro-cue paradigm to manipulate attentional priority in WM, we replicate key behavioral effects from our prior work and address the underlying neural mechanisms in ongoing fMRI analysis.

Keywords: WM; LTM; fMRI; SME; MVPA

Introduction

Which aspects determine if information processed in WM becomes durable in LTM? One possibility is that WM functions as a gateway to LTM and that processes like attentional prioritization or retrieval within WM influence whether information integrates with long-term storage (e.g., Fan & Turk-Browne, 2013; Griffin & Nobre, 2003; McCabe, 2008). In this context, a long-standing question in contemporary WM research is how WM contents survive temporary deprioritization — a phase that may offer a critical window in which WM contents make contact with LTM. Several models propose that deprioritized (i.e., unattended) WM information is offloaded into "activated" LTM (Cowan, 1999; Oberauer, 2002; Beukers et al., 2021). Competing accounts suggest that information could be stored in short-term synaptic ("activity-silent") engrams (Stokes, 2015) or is outsourced to other brain regions (Christophel et al., 2017). A small number of behavioral studies (e.g., Jeanneret et al., 2023; LaRocque et al., 2014; Mao Chao et al., 2023; Strunk et al., 2019) have investigated how WM prioritization affects LTM retention. However, findings have been mixed (Hartshorne & Makovski, 2019): while some studies report enhanced LTM for prioritized items, other work suggests improved LTM for information moved outside the "focus of attention" (Oberauer, 2002; e.g., Rose et al., 2014). Thus, whether and how attentional prioritization affects subsequent LTM remains subject to debate, especially when it comes to linking neural activity during attentional (de)prioritization to LTM outcomes. In our prior behavioral work, we show that overt WM retrieval ("WM testing"), particularly from a deprioritized WM state, benefits LTM (Born & Spitzer, 2024). This suggests that retrieval in WM may function similarly to the well-known "testing effect" in LTM literature (Roediger & Butler, 2011).

Here, we build on these behavioral findings by asking: What are the neural substrates of WM (de)prioritization that predict subsequent LTM retention? To this end, we will examine SMEs to assess how strongly neural activity during a dual retro-cue WM task is predictive of subsequent LTM recall.

Methodology

We conducted an fMRI experiment with n = 26 participants (scanning is ongoing; *n* = 40 planned). The experiment consisted of two stages: (1) a WM task using oriented real-world objects in a dual retro-cue paradigm (Fig. 1a), and (2) a subsequent surprise LTM test for the objects' orientation.

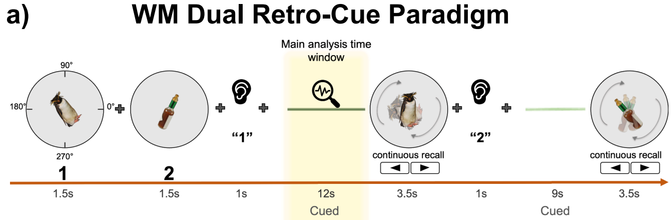
Tasks and procedures. In the WM task (Fig. 1a), participants memorized the orientation of two objects per trial. A retro-cue then indicated which of the two objects would be tested (via a continuous orientation report) after delay 1 (12s). Following delay 1 and WM test, a second retro-cue indicated the same or the other object to be recalled after delay 2 (9s). In this design, during delay 1, one item is prioritized and the other deprioritized (LaRocque et al., 2014). After the WM task, in the subsequent surprise LTM test, participants recalled the orientations of all previously seen WM samples (without scanning; using continuous reports). **Neuroimaging data collection.** We collected whole-brain anatomical and functional data using a 3T Siemens Magnetom TrioTim MRI scanner (isotropic 3 mm^3, Time Repetition = 750 ms)**.** **Planned neuroimaging analysis**. We will focus our analysis on the cued WM maintenance periods (delay 1 & 2) (see Fig. 1a). To capture the fine-grained temporal dynamics of (de)prioritized WM storage, we will estimate the neural activity at each time point and relate these to subsequent LTM performance. We will examine SMEs, an approach that is widely used in fMRI research to link trial-by-trial variability in neural activity during WM processing to subsequent LTM performance (Axmacher et al., 2008; Kim, 2011; Paller & Wagner, 2002). The dependent variable in our SME analysis will be the absolute angular error in LTM recall. We will perform a regression-based SME analysis separately for prioritized vs. deprioritized items in WM. The analysis will specifically focus on regions-of-interest (ROIs) in the occipital cortex, LOC, and IPS previously implicated in visual WM processing (Harrison & Tong, 2009; Huang et al., 2024; Kwak, 2022; Mackey & Curtis, 2017), as well as on medial temporal areas associated with LTM encoding (Kim, 2011; Paller & Wagner, 2002; Xue, 2018). Alongside SME analyses, we will use multivariate pattern analysis (MVPA) (Kriegeskorte et al., 2006) to track and characterize the representation of prioritized and deprioritized WM contents during delay 1 & 2.

Behavioral Results and Conclusions

Figure 1b shows the mean errors (absolute angular difference from the sample orientation; note inverted y-axis) of participants' reports in the WM task. As expected, WM accuracy was highest (i.e., smallest error) for the first-cued items, which had been prioritized during delay 1 [Attended: M = 13.17°, SE = 1.33°]. Accuracy was lower for WM samples cued second in the dual-retro-cue paradigm, particularly for samples that were deprioritized during delay 1 [Unattended: M = 29.25°, SE = 2.84°]. Attended items were recalled with significantly higher accuracy than both attended-repeat [$t(24) = -13.70$, $p < .001$] and unattended items [$t(24) = -8.83$, $p < .001$]. The difference between attended repeat and unattended items was not significant [$t(24) = -1.74$, $p = .094$]. As expected, participants' LTM reports (Fig. 1c) were overall considerably less accurate than their previous WM reports [LTM: M = 60.56°, SE = 3.10°; WM: M = 22.93°, SE = 1.95°; $t(24) = -13.71$, $p < 0.001$]. Critically, preliminary data replicate a WM-"testing" benefit: probed WM items were better remembered in LTM than unprobed ones (M = 71.88°, SE = 3.11°; all $p < 0.01$). Moreover, like in our previous work (double-blind review), 'deprioritized' samples (cued second and recalled only poorly in WM) paradoxically showed the highest LTM accuracy [Attended Repeat: M = 55.38°, SE = 4.08°; Unattended: M = 52.18°, SE = 4.14°] compared to prioritized items [Attended: M = 62.79°, SE = 3.40°].

In sum, this (smaller) fMRI sample replicates our key behavioral finding from a large online sample (n = 450): a robust 'WM-testing effect', particularly for content that had been temporarily deprioritized in WM. The results complement existing LTM-"testing" literature by suggesting a similar retrieval practice benefit for overtly retrieved WM content. In our (ongoing) fMRI analysis, we anticipate replicating well-established SMEs, linking neural activity during encoding to subsequent memory performance.

a) WM Dual Retro-Cue Paradigm



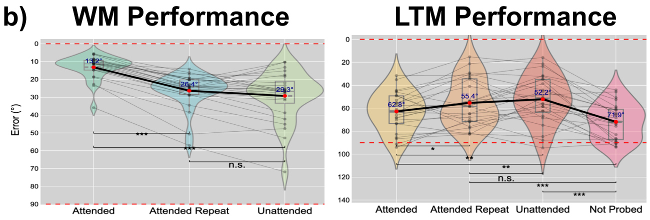


Figure 1: Behavioral paradigm overview and behavioral WM and LTM results. *a)* Dual-retro-cue WM task. *b)* WM Performance (left) and LTM performance (right). Red dots and dark blue numbers represent means; grey dots, individual participants. Violin outlines illustrate the distribution over participants using kernel density estimation. Dashed horizontal lines mark ceiling (0°) and chance-level performance (90°).

Acknowledgments

This research was supported by European Research Council Consolidator Grant ERC-2020-COG101000972 (B.S.). The funders had no role in the study design, data collection, and analysis, decision to prepare this extended abstract as CCN submission.

References

Axmacher, N., Schmitz, D. P., Weinreich, I., Elger, C. E., & Fell, J. (2008). Interaction of Working Memory and Long-Term Memory in the Medial Temporal Lobe. *Cerebral Cortex*, *18*(12), 2868–2878. https://doi.org/10.1093/cercor/bhn045

Beukers, A. O., Buschman, T. J., Cohen, J. D., & Norman, K. A. (2021). Is activity silent working memory simply episodic memory? *Trends in Cognitive Sciences*, *25*(4), 284–293.

Born, F., & Spitzer, B. (2024). Long-Term Effects of Working Memory Retrieval From Prioritized and Deprioritized States. https://doi.org/10.1101/2024.10.08.617145

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J.-D. (2017). The distributed nature of working memory. *Trends in Cognitive Sciences*, *21*(2), 111–124.

Cowan, N. (1999). An embedded-processes model of working memory. *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, *20*(506), 1013–1019.

Fan, J. E., & Turk-Browne, N. B. (2013). Internal attention to features in visual short-term memory guides object learning. *Cognition*, *129*(2), 292–308.

Griffin, I. C., & Nobre, A. C. (2003). Orienting attention to locations in internal representations. *Journal of Cognitive Neuroscience*, *15*(8), 1176–1194.

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635.

Hartshorne, J. K., & Makovski, T. (2019). The effect of working memory maintenance on long-term memory. *Memory & Cognition*, *47*, 749–763.

Huang, J., Wang, T., Dai, W., Li, Y., Yang, Y., Zhang, Y., Wu, Y., Zhou, T., & Xing, D. (2024). Neuronal representation of visual working memory content in the primate primary visual cortex. *Science Advances*, *10*(24), eadk3953. https://doi.org/10.1126/sciadv.adk3953

Jeanneret, S., Bartsch, L. M., & Vergauwe, E. (2023). To be or not to be relevant: Comparing short- and long-term consequences across working memory prioritization procedures. *Attention, Perception, & Psychophysics*, *85*(5), 1486–1498. https://doi.org/10.3758/s13414-023-02706-4

Kim, H. (2011). Neural activity that predicts subsequent memory and forgetting: A meta-analysis of 74 fMRI studies. *NeuroImage*, *54*(3), 2446–2461. https://doi.org/10.1016/j.neuroimage.2010.09.045

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, *103*(10), 3863–3868. https://doi.org/10.1073/pnas.0600244103

Kwak, Y. (2022). *Unveiling the abstract format of working memory representations*. https://doi.org/10.17605/OSF.IO/T6B95

LaRocque, J. J., Lewis-Peacock, J. A., & Postle, B. R. (2014). Multiple neural states of representation in short-term memory? It's a matter of attention. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00005

Mackey, W. E., & Curtis, C. E. (2017). Distinct contributions by frontal and parietal cortices support working memory. *Scientific Reports*, *7*(1), 6188. https://doi.org/10.1038/s41598-017-06293-x

Mao Chao, C., Xu, C., Loaiza, V., & Rose, N. S. (2023). Are latent working memory items retrieved from long-term memory? *Quarterly Journal of Experimental Psychology*, 17470218231217723. https://doi.org/10.1177/17470218231217723

McCabe, D. P. (2008). The role of covert retrieval in working memory span tasks: Evidence from delayed recall tests. *Journal of Memory and Language*, *58*(2), 480–494. https://doi.org/10.1016/j.jml.2007.04.004

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention.

Journal of Experimental Psychology: Learning, Memory, and Cognition, *28*(3), 411–421. https://doi.org/10.1037/0278-7393.28.3.411

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, *6*(2), 93–102. https://doi.org/10.1016/S1364-6613(00)01845-3

Reaves, S., Strunk, J., Phillips, S., Verhaeghen, P., & Duarte, A. (2016). The lasting memory enhancements of retrospective attention. *Brain Research*, *1642*, 226–237.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27.

Rose, N. S., Buchsbaum, B. R., & Craik, F. I. (2014). Short-term retention of a single word relies on retrieval from long-term memory when both rehearsal and refreshing are disrupted. *Memory & Cognition*, *42*(5), 689–700.

Stokes, M. G. (2015). ‘Activity-silent’ working memory in prefrontal cortex: A dynamic coding framework. *Trends in Cognitive Sciences*, *19*(7), 394–405. https://doi.org/10.1016/j.tics.2015.05.004

Xue, G. (2018). The neural representations underlying human episodic memory. *Trends in Cognitive Sciences*, *22*(6), 544–561.