CellPress

Trends in Cognitive Sciences

Spotlight

Rapid connectivity modulations unify long-term and working memory

Eelke Spaak[1,*], and Michael J. Wolff[2]

Panichello *et al.* recently demonstrated that working memory (WM) information can be maintained without active neural firing. Instead, it is stored in rapidly modulating neural connectivity patterns. This validates the activity-silent model of WM, and unifies the mechanisms of long-term memory (LTM) and WM. Here, we highlight the ramifications of these findings.

WM is the ability to hold information 'in mind' over short time spans, and is critical to many forms of cognition, from remembering a phone number to holding a conversation. Given this critical importance, and the fact that WM impairments are central in many neurological disorders, understanding the neural underpinnings of WM is a key objective in cognitive neuroscience. Originally driven by the discovery of WM-selective neurons in primate prefrontal cortex (PFC), a long-held, still dominant, view of WM maintenance is that it depends on persistent neuronal activity during memory delay periods [1]. This is in contrast to models of LTM: these propose that LTM is maintained in the patterns of (synaptic) connectivity among neurons [2].

Several researchers have challenged the persistent-firing view of WM and instead proposed that WM is maintained in 'activity-silent' neural ensembles in PFC, rapidly modulating effective (synaptic) connectivity patterns to encode memoranda [3,4]. This model has clear theoretical advantages over WM models that rely on persistent neural firing, such as reduced metabolic costs and increased robustness to distractors. Importantly, it also entails a particularly efficient WM readout: if WM is stored in the connectivity patterns of a network, then the processing of future input by that network necessarily depends on its past stimulation history [4]. This abolishes the need for a separate mechanism comparing stored memoranda with new input, and links naturally to a now-emerging understanding of memory as directed at the future, rather than the past [5]. Overall, if connectivity plays a key role in WM, then its neural underpinnings would be more similar to those of LTM than previously thought.

Although indirect evidence for this connectivity-based view of WM has been reported previously (e.g., [6]), a direct test has proven challenging. A recent breakthrough [7] enables simultaneous recordings of hundreds of neurons spaced closely together. Leveraging this novel technology, Panichello *et al.* recently demonstrated that WM is indeed maintained across activity-silent periods, and that rapid modulations of connectivity play a crucial role [8].

Panichello *et al.* started by replicating the classic finding that population activity during WM delays shows continuous decoding of the memorandum. Importantly, once one goes beyond the trial average, single-trial data show a far from continuous view. Instead, the population alternates between 'On' states, with clear WM-specific spiking activity, and 'Off' states, with a clear absence of such. These state differences could not be explained by differences in background noise or coding scheme. Critically, the On/Off states are remarkably robust, and occur in coordinated fashion across the local population. Silent periods of some neurons are thus not bridged by active periods of other neurons.

Further leveraging the massively parallel nature of their recordings, the authors proceeded to analyze connectivity between neurons. They found high numbers of connected pairs (i.e., pairs of neurons showing a short-latency peak in their cross-correlogram). The probability of two neurons being connected decreased with distance, as expected from anatomy, lending support to the robustness of this pipeline. This could explain why previous studies have struggled to test connectivity modulations: electrode contacts are typically spaced further apart than in the probes used here, greatly reducing the chance of identifying neurons that are connected at all, let alone any modulation of such connectivity.

Critically, Panichello *et al.* found that patterns of connectivity differed among different memory conditions, particularly during the Off ('activity-silent') states. In other words, they reported the first robust evidence that WM content is reflected in rapidly modulating (synaptic) connectivity in PFC.

Finally, Panichello *et al.* reported that neuron pairs with similar memory-selective spiking responses during presentation of the to-be-remembered stimulus are more likely to have a memory-selective connectivity modulation during the delay period: neurons that fire together during encoding, wire together for maintenance. This is akin to Hebbian plasticity leading to novel synaptic connections in LTM (Figure 1A). However, given the timescale and flexibility necessary for WM, it is almost certainly existing connections that rapidly modulate their efficacy for WM maintenance, allowing a quick return to a baseline, WM-neutral, state (Figure 1B).

We posit that connectivity changes for WM are best thought of not as a short-term buffer for the recent past, but rather as an efficient, future-oriented readout mechanism [4]. In line with this, different (silently maintained) WM contents result in systematically different neural responses to

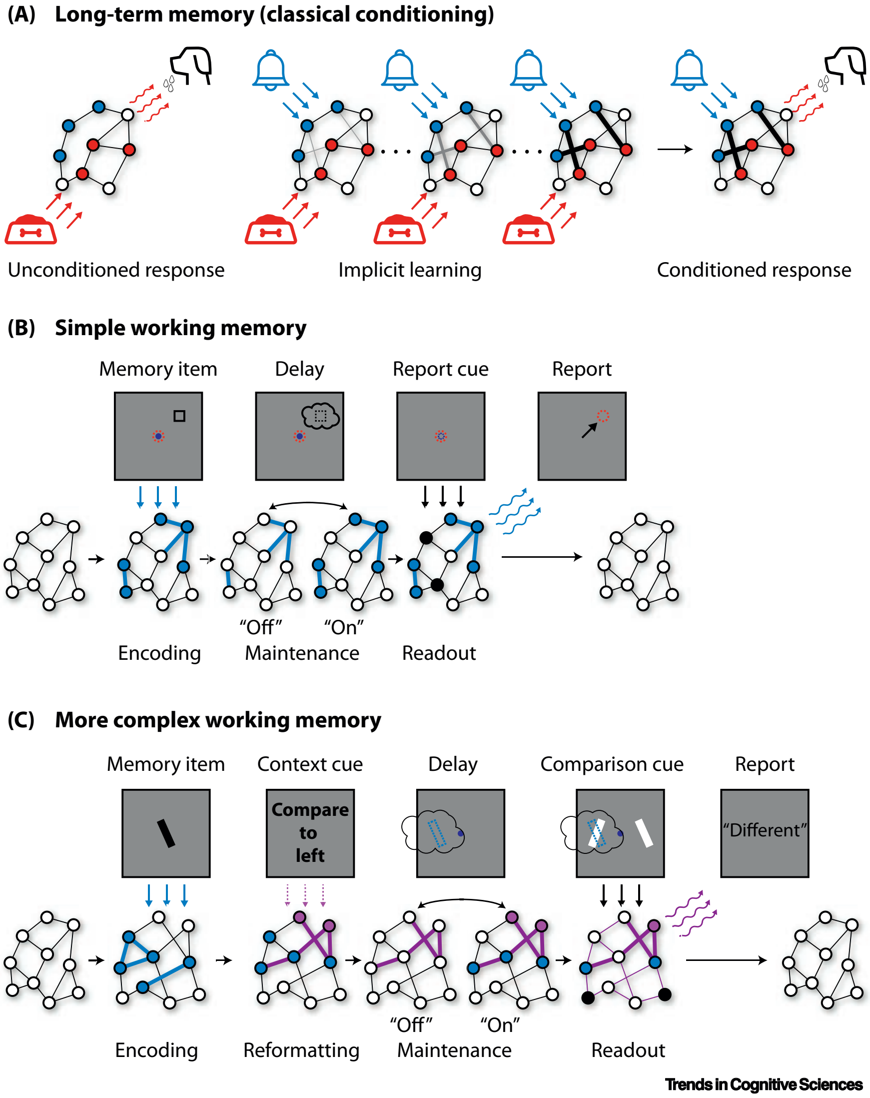

Figure 1. Neural similarities and differences between long-term memory (LTM) and working memory (WM). (A) Conditioning results in new neural connections. The network reconfigures to produce a conditioned response (salivation) to a previously neutral cue (bell). (B) In Panichello *et al.*[8], neurons jointly selective to the same location modulate their connections during encoding to silently maintain the memorandum during the delay. The report cue results in WM-specific readout. (C) Hypothetical context-dependent WM task. A context cue evokes neural activity that reformats the memorandum, triggering novel connectivity modulations. We predict that this reduces the overlap between jointly selective neurons during encoding and pairs with modulated connections during maintenance.

subsequent neutral external stimulation [9]. When the stimulation is not neutral, but task relevant, the network is configured to respond in alignment with the expected task demands. In Panichello *et al.*'s case, the disappearance of the fixation dot triggers an active neural response in the memory-configured WM network that leads to the appropriate behavioral response: an eye movement to the remembered location (Figure 1B).

That brings us to our final takeaway. Panichello *et al.* found high neural similarity between spiking activity in the cue period and the delay period [8]. This may be a consequence of the high similarity between the cue itself and the behavioral goal (i.e., both are the same spatial location). More complex WM tasks result in more complex neural dynamics [10]. This possibly reflects the more sophisticated demands that such tasks place on the appropriate connectivity pattern for optimal readout [4]. We speculate that complex WM tasks, requiring a reformatting of the memory item to optimally guide behavior, will result in reduced overlap between jointly selective neurons during encoding and pairs with modulated connections (Figure 1C). This means that the more complex the task, the less 'fire together during encoding, wire together for maintenance' will hold. The delay period should still show significant connectivity modulation, only in different subsets of neurons.

We anticipate that understanding this rich interplay between active neuronal activity and modulations of effective connectivity, now shown to be relevant at all timescales, will be a key and fruitful research goal in the years to come, not only for the cognitive neuroscience of memory, but for the field in general.

Acknowledgments

Dedicated to the memory of Mark Stokes (1976–2023): no longer actively with us, but having left a persistent trace.

Declaration of interests

The authors declare no competing interests.

[1]Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, The Netherlands
[2]Ernst Strüngmann Institute for Neuroscience, Frankfurt am Main, Germany

*Correspondence:
eelke.spaak@donders.ru.nl (E. Spaak).

https://doi.org/10.1016/j.tics.2025.02.006

References

1. Sreenivasan, K.K. and D'Esposito, M. (2019) The what, where and how of delay activity. *Nat. Rev. Neurosci.* 20, 466–481

2. Martin, S.J. *et al.* (2000) Synaptic plasticity and memory: an evaluation of the hypothesis. *Annu. Rev. Neurosci.* 23, 649–711
3. Mongillo, G. *et al.* (2008) Synaptic theory of working memory. *Science* 319, 1543–1546
4. Stokes, M.G. (2015) 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn. Sci.* 19, 394–405
5. van Ede, F. and Nobre, A.C. (2023) Turning attention inside out: how working memory serves behavior. *Annu. Rev. Psychol.* 74, 137–165
6. Fujisawa, S. *et al.* (2008) Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nat. Neurosci.* 11, 823–833
7. Steinmetz, N.A. *et al.* (2021) Neuropixels 2.0: a miniaturized high-density probe for stable, long-term brain recordings. *Science* 372, eabf4588
8. Panichello, M.F. *et al.* (2024) Intermittent rate coding and cue-specific ensembles support working memory. *Nature* 636, 422–429
9. Wolff, M.J. *et al.* (2017) Dynamic hidden states underlying working-memory-guided behavior. *Nat. Neurosci.* 20, 864–871
10. Spaak, E. *et al.* (2017) Stable and dynamic coding for working memory in primate prefrontal cortex. *J. Neurosci.* 37, 6503–6516