

## Opinion

## Is Activity Silent Working Memory Simply Episodic Memory?

Andre O. Beukers,<sup>1,\*</sup> Timothy J. Buschman,<sup>1,2</sup> Jonathan D. Cohen,<sup>1,2,3</sup> and Kenneth A. Norman<sup>1,2,3</sup>

**Working memory (WM) maintains task-relevant information in a state ready for processing. While traditional theories assume that sustained neuronal activity is responsible for WM, the Activity Silent WM (ASWM) account proposes that maintenance can also be supported by short-term synaptic weight changes. Here, we argue that the evidence for ASWM can be explained more parsimoniously by the involvement of episodic memory (EM) in WM tasks. Like ASWM, EM relies on rapid synaptic modification that is also activity silent; however, while ASWM posits transient synaptic modifications, EM traces persist over longer time periods. We discuss how, despite this difference, well-established EM mechanisms can account for the key findings attributed to ASWM, and describe predictions of this account.**

**Activity Silent Maintenance and EM**

**WM** (see [Glossary](#)) is the ability to maintain and manipulate task-relevant information [1]. While WM has traditionally been thought to rely on the **active maintenance** of neural representations [2–6], recent studies have shown that neural activity often returns to baseline during the delay periods, suggesting that information is not always actively maintained [7–10]. This **activity silent memory** puzzle motivated the **ASWM** account, which supplements active maintenance in WM with a mechanism based on short-term synaptic weight changes [11–16]. Here, we question whether observations of activity silent memory warrant the postulation of a new memory mechanism. ASWM implements a function (rapid learning via synaptic weight changes) that is already supported by an established memory system; namely, **EM**. In the first part of this opinion article, we reinterpret the neurophysiological evidence for ASWM as evidence for EM's involvement in **WM tasks**, with a specific focus on the role of structures critically implicated in EM [i.e., hippocampus (HPC), medial temporal lobe (MTL)].

Importantly, evidence for the ASWM account comes from findings of activity silent memory on tasks assumed to specifically target WM [12,14,16]. These WM tasks are characterized by a small stimulus set that is reused over multiple trials with short maintenance durations. Under these circumstances, persisting traces of similar or identical items should accumulate in EM and thereby impair participants' ability to distinguish the presence of stimuli in the current trial from their appearance in previous trials. That is, WM tasks are specifically designed to prevent EM involvement by posing a high risk of **proactive interference (PI)**. Note that, in contrast to EM traces that persist over long time periods, extant ASWM models assume that traces are transient, either because they degrade if not refreshed [10,14] or because they are overwritten by new information [17,18]. Thus, in these models, the transience of WM traces avoids the issue of PI (discussed below). A key motivation for positing the existence of ASWM is the idea that activity silent memory can occur in tasks that are specifically thought to exclude EM; if EM is truly excluded, some other mechanism (ASWM) must be posited to explain these results. In the second part of this paper we challenge the idea that EM cannot support WM task

**Highlights**

The evidence for the Activity Silent Working Memory (ASWM) model can be interpreted as evidence of the involvement of episodic memory (EM) in working memory (WM) tasks.

EM is subject to proactive interference (PI), which has been assumed to limit its engagement in WM tasks; however, context representations actively maintained in WM can be used to mitigate PI, thereby allowing EM to be effectively used in WM tasks.

The hypothesis that EM is involved in WM tasks leads to testable predictions about how the neural substrates of EM (e.g., the hippocampus) should be engaged in these tasks.

This hypothesis also implies that, to predict behavioral memory performance in these tasks, we need to measure the structure of context representations maintained in WM and how these change over time (e.g., larger context shifts should lead to less PI).

<sup>1</sup>Department of Psychology, Princeton University, Princeton, NJ, USA  
<sup>2</sup>Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA  
<sup>3</sup>Co-last authors

\*Correspondence:  
[abeukers@princeton.edu](mailto:abeukers@princeton.edu) (A.O. Beukers).

performance. We discuss how **context representations** can be used to distinguish and retrieve appropriate EM traces in a manner that prevents crippling levels of PI and allows EM to contribute. We describe testable predictions that arise from this account regarding the similarity structure of actively maintained context representations in WM and how this shapes behavioral memory performance.

## What Is the Neural Substrate of Activity Silent Memory?

### Activity Silent Memory

Observations that item-specific neural activity returns to baseline during the maintenance period of a WM task [19] have led researchers to challenge the classic assumption that active maintenance is the neural substrate of WM [3,4] (although see [2]). Dovetailing with psychological accounts of WM [11,20,21], the ASWM account distinguishes between attended and unattended memory items: While the attended memory item is actively maintained, unattended memory items can be silently maintained. Evidence for this distinction comes from multiple paradigms, most prominently the dual retrocue paradigm [7,8,10]. In this paradigm, two stimuli are presented at the start of each trial; next, participants are given a cue indicating which of the two stimuli will be tested, followed (after a delay) by a probe that tests memory for the cued stimulus; and finally, participants are given a second cue, followed by a second probe. Interestingly, which of the two items can be decoded from neural activity changes over the trial: At stimulus presentation, both stimuli are easily decodable, but during the delay period after the first cue, only the cued stimulus can be decoded, while decoding of the other stimulus drops to baseline. However, information about the second item is not completely gone: if the participant is subsequently cued to report the other stimulus, the previously unrepresented stimulus can now be decoded, and thus must have been retained [7,8,10].

The ASWM interpretation of these results is that when neural activity ‘disappears’, this is because the unattended memory is encoded in rapidly formed, but short-lasting, weight modifications that can be read out from the network when needed [11,12,16]. However, this pattern of stimulus decoding is also consistent with storage in EM, defined as the ability to retain briefly presented information via rapidly formed but longer-lasting weight modifications, and its later reinstatement to an active form when needed. While the HPC and surrounding MTL have long been known to be critical for EM functioning [22], more recent work suggests that the MTL may also be important in WM tasks. For example, Olson and colleagues [23,24] have shown that when memory items are not conducive to verbal rehearsal (e.g., visual materials), patients with MTL damage are impaired in tasks with delays as short as 4 s. More recent work suggests that visual WM may be impaired in such patients with delays as short as 1 s [25]. Recent neuroimaging evidence is also consistent with the use of EM in these tasks, even under conditions traditionally assumed to favor WM [26].

This growing literature showing signatures of EM involvement in WM tasks substantiates proposals that WM and EM work in concert during these tasks [27–29]. To test for the role of EM in activity silent memory, future studies should investigate the role of the HPC and MTL in the dual retrocue paradigm. For example, does disruption of processing in the HPC and/or surrounding MTL regions disrupt activity silent memory? An important point to keep in mind here is that, in this opinion article, we define memory systems functionally, the critical distinction being whether traces encoded in activity silent memory are durable (as suggested by EM) or transient (as suggested by ASWM). Thus, while a disruption of activity silent memory associated with damage to the HPC and/or MTL would be informative, failure to find this would leave open the possibility that other preserved structures are capable of supporting performance through durable traces; if this were the case, we would attribute this to EM according to the functional definition spelled out here.

### Glossary

**Active maintenance:** process that retains information in patterns of neural firing; a key signature of this process is that retained information at each time point can be decoded from neural activity.

**Activity silent memory:** retention of a memory for a period during which any neural activity associated with the memory becomes indistinguishable from baseline levels (measured outside the retention period).

**Activity Silent Working Memory (ASWM):** an account of activity silent memory positing that memories can be retained through short-term synaptic plasticity (i.e., rapidly forming, transient associations) in addition to active maintenance in WM.

**Context representation:** neural activity that is not directly related to the sensory and semantic features of currently attended items; posited to have an important role in organizing memory retrieval.

**Episodic memory (EM):** memory system that can support the formation of arbitrary novel associations (e.g., item to context) and the recall of these associations over an extended period of time, after a single study exposure; neurally instantiated as rapid long-term plasticity; thought to depend on the HPC.

**Position code:** a context representation that carries information about the serial position of an item.

**Proactive interference (PI):** situation in which previously formed memories impair present task performance; arises when many similar memories coexist.

**Temporally drifting context:** a context representation that changes smoothly (i.e., drifts) over time and thus carries information about the passage of time.

**Trial tag:** a context representation that changes abruptly at trial boundaries and thus serves to distinguish memories formed in different trials.

**WM task:** memory task designed to isolate WM processes; characterized by short delay periods and requiring flexible updating of contents maintained in memory.

**Working memory (WM):** a flexible memory system that retains information over a period of seconds to minutes, beyond immediate sensory inputs. The prefrontal cortex is thought to be a key neural substrate of WM.

### Network Perturbation Increases Memory Decodability

In [16] it is reasoned that, if activity silent WM is maintained by synaptic weight changes, these changes should alter how the network responds to a stimulus in a manner that is specific to the memories being stored. In support of this view, one study [8] showed that a transcranial magnetic stimulation (TMS) pulse applied to the scalp during the activity silent maintenance period produced memory-specific neural signatures. Similarly, another study [10] showed that briefly flashing a task-irrelevant stimulus during activity silent maintenance produced different neural responses for different maintained stimuli. While perturbation studies show that non-specific stimulation drives different response patterns, and thereby provide strong support for the existence of weight-based traces, they do not speak to the neural substrate of these latent traces or their duration. While ASWM assumes that the traces reside in short-lasting weight changes made to the cortex, the existing evidence is also consistent with the possibility that the effects of perturbations are mediated by longer-lasting traces associated with EM mechanisms (e.g., in the HPC). TMS stimulation has been shown to activate the cortical–hippocampal network [30] and to enhance HPC-based memory [31]. Similarly, visual stimulation, as in [10], can trigger neural firing in the visual cortex that propagates up the visual hierarchy to the HPC, eliciting EM retrieval. Future studies could test this hypothesis by investigating the effects that perturbation manipulations have on HPC activity, specifically looking for signatures of neural plasticity, increased HPC recruitment, cortical–hippocampal interactions, and/or signatures of EM reinstatement (e.g., pattern completion).

### WM Maintenance Is Accompanied by Activity Bursts

One of the critical differences between WM and EM is the timespan of trace persistence. It is widely assumed that activity that supports WM traces degrades, decays, or is rapidly overwritten and that refreshing and rehearsal mechanisms are necessary to maintain WM traces (e.g., through an articulatory loop) [1]. ASWM proposes that this is complemented by mechanisms to sustain traces using short-term plasticity that, like sustained activity, is transient but can be refreshed. For example, recent studies have suggested that brief bursts of high-frequency oscillatory activity (i.e., gamma bursts) in the frontal cortex serve to prevent ASWM traces from degrading [14]. In support of this idea, it has been shown that bursting in the gamma range increases with WM load and is selective to the neurons that encode the maintained stimulus representation, and that the burst rate increases towards the end of the delay period [16]. While these results have been interpreted as evidence for ASWM refreshing [14,32–34], they are also consistent with the possibility that gamma is a neural signature of the reinstatement of EM traces into WM. Likewise, the finding that gamma suppression is related to reduced levels of stimulus information has been interpreted as ASWM forgetting [35], but this could alternatively be explained in terms of gamma being lower when EM reinstatement is being inhibited. According to this alternative account, gamma bursts play a role in WM–EM interactions and are thus predicted to be synchronized between cortical sites carrying WM information (e.g., the PFC) and regions supporting EM function (e.g., the MTL). Gamma has been shown to mediate communication between the PFC and HPC [36] and HPC gamma has been found to be sensitive to memory load in the context of WM tasks [37]. This sensitivity to memory load (in both the HPC and PFC) suggests a potential link between PFC gamma [32] and HPC gamma [37]. To test the proposal that gamma bursts reflect reinstatement of EM traces into WM, future neurophysiological studies could use electrophysiological recording techniques to investigate whether and (if so) how HPC and PFC gamma support activity silent memory.

### The Role of EM in WM Tasks

The first four rows of Table 1 review how phenomena cited in support of ASWM are also consistent with reliance on EM (see also [28]). However, as mentioned in the introduction, the durability

Table 1. Observed Phenomena Along with ASWM and EM Interpretation

Phenomenon	ASWM explanation	EM explanation
Activity silent maintenance	Short-term plasticity, likely in PFC	Long-term plasticity, likely in HPC
Network perturbation increases decodability	Different activity flows due to different network weights	Promotes EM reinstatement by recruiting HPC
WM maintenance is accompanied by activity bursts	Activity bursts refresh short-term traces by activating representations	Activity bursts are a signature of retrieval from EM into WM
Increased gamma burst rate towards end of delay period	Recoding from silent to active WM trace	Increasing probability that memory is retrieved from EM into WM
Flexible forgetting	Traces are transient	Decreased accessibility by context shift

of EM traces poses a computational challenge: If participants are relying on durable EM traces, there is the possibility of a crippling buildup in PI as these traces accumulate across trials. PI is not an issue for ASWM, since traces from past trials are assumed to not persist. One proposed solution to mitigate the PI of accumulating EM traces is to actively delete old traces [38]. While some preliminary neural evidence has been obtained for this kind of ‘active deletion’ mechanism [39], we argue here for an alternative; namely, that the management of PI can be accounted for by existing theories of EM, without having to resort to active deletion of activity silent traces. In short, we propose that the encoding of context information along with stimuli in EM provides a means of overcoming PI. Below, we outline an interactive WMEM account (Figure 1, Key Figure) that describes how context representations maintained in WM implement a form of cognitive control that can be used to bias EM retrieval, thereby mitigating PI and allowing EM to be used as an activity-silent source of memory in WM tasks (see [28] for a closely related view). As emphasized below, WMEM is not a new theory as such, but rather an account that brings together well-established ideas from both the EM and the cognitive control literature to explain how EM is engaged and controlled in WM tasks (for related ideas about WM–EM interactions, see [21,40–44]).

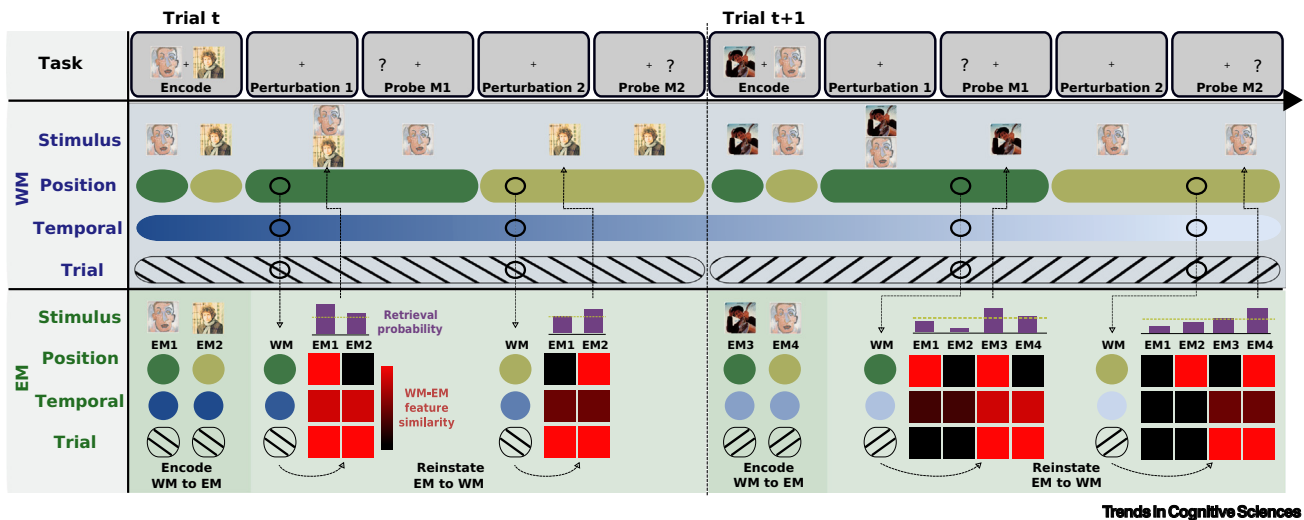
#### WMEM Account

WMEM has three core principles: First, WM is a system that actively maintains a collection of features of currently attended items (e.g., size, color, shape, orientation, meaning) and information not directly related to these features; here, we use the term ‘context representations’ as a catch-all to refer to the latter type of information. While context representations can encompass many heterogeneous types of information, we focus here on three types of context representations: **temporally drifting context** that changes gradually within and across trials; **trial tags** that shift abruptly at trial boundaries; and **position codes** that represent specific serial positions. The second principle of WMEM is that EM traces are encoded rapidly and persist over extended periods of time; EM traces bind together all of the features that are active in WM at the time of encoding, including features of the currently attended stimulus [3,45–48] as well as any context representations that happen to be currently active. Neurobiologically, this means that on each trial of a memory task, information about the currently activated pattern of cortical activity is encoded as a synaptic trace, with hippocampal long-term potentiation likely to play an important role in this process (for a review see [49]). Third, according to the encoding specificity principle [50], retrieval from EM is a function of the similarity between the (silently) stored EM trace and the cues that are (actively) represented at retrieval (Figure 1). If the actively represented information in WM matches the silently stored EM trace, that EM trace is reinstated in the cortex [51,52].

The aforementioned principles are fully in line with classic context-based theories of EM [53–58]. According to these theories, EM encoding involves the binding together of a representation of the stimulus with a context representation that gradually drifts over time. Later, those context

## Key Figure

## Interactive Working Memory Episodic Memory (WMEM) Account



**Figure 1.** Predicted contents of and interactions between WM and EM over two trials of the Wolff *et al.* [10] dual retrocue paradigm involving visual perturbations (see text for additional paradigm details). On each trial, participants are given two memory stimuli, which they must report in turn. Before the memories are reported, a visual perturbation (unrelated to the memory stimuli) is presented that triggers a neural response, from which information about the contents of memory can be decoded. On the top is a representation of two consecutive trials (trial  $t$  and trial  $t + 1$ ; time along the x-axis). Contents of WM are represented by four rows corresponding to the stimulus and three different types of context (position codes, temporally drifting context, and trial tags). Note that position codes are discrete and reoccur, temporally drifting context changes gradually, and trial tags shift to new values at trial boundaries. Below WM is a representation of the contents of EM. During the encoding period, EM stores the currently active contents of WM. Retrieval from EM occurs as follows. First, the current WM state (downward-pointing arrow) is compared with EM traces (rightward curved arrow) on a feature-by-feature basis (heatmaps; red–black indicates high–low similarity for each feature). Next, the aggregate feature similarity is combined into a retrieval probability value (purple bar graph; horizontal yellow line indicates retrieval threshold). Finally, the EM traces that exceed a retrieval threshold are reinstated into WM (upward vertical arrows). On trial  $t$ , we show how perturbations applied at different timepoints trigger EM retrieval differently; on trial  $t + 1$ , we show how proactive interference from trial  $t$  is prevented and how position codes specifically target different memories for intentional retrieval when probes appear (see text for more details on the contributions of the different types of context).

representations can be used to cue recall of the stimulus; stimulus traces that were bound to those elements of context (or similar ones) will be more likely to be retrieved than traces that were linked to other contexts (Box 1). In these models, as in the WMEM framework, EM traces are assumed to endure over time. These models explain the recency effect (i.e., the finding that recent memories are easier to retrieve than more distant memories) not in terms of any ‘fading’ of memories, but rather in terms of differential cue–trace match: because of contextual drift, the currently active context at test matches the context that was bound to recent memories better than it matches the context bound to more distant memories. The WMEM framework joins these ideas about context and EM with ideas from the cognitive control literature, positing that control operates by biasing processing in favor of task-relevant representations [59–61]. Later, we discuss how different forms of context (that vary both between and within trials) implement a biasing control on EM [62] that mitigates PI, thereby allowing EM to inform behavior during WM tasks.

#### Trial Tags Reduce between-Trial PI

The most efficient way to avoid PI from EM traces formed on previous trials would be to use a context representation that is perfectly stable within a trial but dramatically shifts across trial boundaries. There is growing evidence for the existence of shifting representations, from both animal neurophysiology studies [63–65] and human neuroimaging studies [66–69]. By being

### Box 1. Shifting Context Representations and EM Organization

The idea that actively maintained context representations can organize EM has a long history, dating back to work positing a role for random context drift in the organization of memories [58,84]. More recent models have elaborated on this view, replacing the idea of random drift with the idea that context drift arises via the integration of features of recently encountered stimuli [53,56]; neurobiologically, this integration process is compatible with data showing stimulus-specific coding in hippocampal time cells [72,73] and conjunctive what/when codes in the frontal cortex [75]. Recent models of how context influences EM also posit that items can trigger the reinstatement of previous contextual states associated with those items ('jumping back in time' [53,56,85]; for human neural data that support this claim, see [86,87]).

Alongside the evidence for drifting mental context, there is growing appreciation for the idea that context representations can abruptly shift when people's inferences about the underlying situation or 'latent cause' change (for reviews see [67,88]). By serving to separate EMs on either side of the shift, these shifts can modulate the retrievability of subsets of memories [89]. For example, a recent fMRI study [90] found that when participants were instructed to forget (versus remember) a list they were just presented, there was a greater mental context shift (measured using fMRI) and the size of the context shift correlated with memory retrievability across participants. In parallel with the above developments in the human EM literature, researchers (largely, but not entirely, focusing on animal data) have worked to characterize a complex 'zoo' of cognitive codes that arise during learning and task performance [91], many of which have the 'shifting' property noted above. For example, in [65] it was found that there are neural codes in the HPC that are stable while a rat completes a lap around a circular maze and shift abruptly when a new lap starts. In [63] it was shown that shifting hippocampal representations are related to lower levels of PI. We speculate that all of the above types of context codes could be leveraged to structure the use of EM during WM tasks.

stable within a trial, these 'trial tag' representations encourage the retrieval of memories formed within that trial; by shifting at boundaries, trial tags discourage the retrieval of memories formed on previous trials (Figure 1). A key prediction arising from WMEM is that the separation in trial tag representations should be related to EM retrieval and thereby performance; better neural separation between trials should lead to lower PI and (through this) better memory performance [63]. Another implication of this view is that, if a trial tag representation is corrupted in WM within a trial, this will impair retrieval.

### Within-Trial Flexibility Is Accomplished by Position Codes

Earlier we described trial tags that can help to differentiate EM traces on a trial-by-trial basis. However, WM tasks often require flexible updating of memories not only between trials but also within a trial. For example, in [10] a version of the dual retrocue task was conducted in which two memory items were presented on each trial, one on the left side of the screen [left memory (LM)] and one on the right side of the screen [right memory (RM)]; instead of cueing each in turn as in the standard dual retrocue, participants were taught that they would have to respond based on LM for the first probe (P1) and based on RM for the second probe (P2). Critically, when a perturbation was applied before P1, both LM and RM were decoded from neural activity. However, when a perturbation was applied between P1 and P2 (i.e., right after P1 but before P2), only RM was decoded. This pattern of results shows that information retrieved from activity silent memory can be rapidly and flexibly updated within a trial. ASWM can potentially explain this in terms of selective refreshing of the trace of RM and letting the trace LM degrade; however, to work properly, this requires some way of specifying which traces should be refreshed.

By contrast, WMEM accounts for flexible activity silent memory updating using context representations. In addition to trial tags, which vary from trial to trial, other components of context vary within a trial (Box 1). For example, recent studies have shown that representations of serial position (i.e., position codes; Figure 1) can be detected in neural activity [69–71]. Given that these position codes are actively represented, they could be encoded as features in EM traces. If different position codes are maintained along with LM and RM on each trial, these different codes could serve to distinguish between EM traces (Figure 1). For example, when intentionally retrieving one memory or another, reinstatement of the corresponding position code would bias retrieval of the

appropriate EM trace. Also, returning to the perturbation findings described earlier, a combination of position codes and temporal drift could account for the observed pattern of results. When the first perturbation is applied, the slowly drifting context representation is still similar enough to its state at encoding to push both traces above the retrieval threshold; this explains why the first perturbation drives the retrieval of LM and RM. As the trial goes on, and the second perturbation is eventually applied, the temporal drift no longer matches the encoding period of either EM trace. However, because of the position code maintained in WM, the perturbation still encourages the retrieval of RM (Figure 1). As with trial tags, WMEM predicts that this pattern of temporal drift and position codes should be detectable from neural activity; there is already abundant evidence for temporally drifting neural codes in other paradigms [72–75], although their role in paradigms like the one used by [10] still needs to be established empirically.

### Concluding Remarks

In this opinion article, we argue that the involvement of EM in WM tasks can account for the phenomena associated with activity silent memories. Critically, we propose an interactive account that combines existing theories of WM [27] and EM [54] to explain these phenomena, an approach that is arguably more parsimonious than postulating a new ASWM system. While existing evidence indicates that transient synaptic changes do occur in the PFC [76,77], it is not yet known whether these transient changes are responsible for the phenomena that ASWM seeks to explain. To be clear, we are not arguing that the account we propose is mutually exclusive with ASWM. Our goal is to articulate a WMEM account that motivates future experimental work to test predictions of each account and determine which – or whether both – are supported by the data. To that end, we have emphasized the role of context representations maintained in WM that serve to implement a biasing control on EM, mitigating PI and thus allowing EM traces to inform performance on WM tasks (see Outstanding Questions for an overview of predictions).

One might reasonably be concerned that, in placing such a strong emphasis on the use of actively maintained context to retrieve (latent) stimulus representations, we are placing an unreasonable load on limited-capacity WM systems: if participants in dual retrocue studies are unable or unwilling to actively maintain the two stimuli over time, how are participants able to actively maintain all of the context representations shown in Figure 1? Our answer here is that, from an information-theoretic perspective, context representations are ‘cheap’ to maintain (in terms of the number of bits required). Temporally drifting context does not need to hold any recoverable information, it just needs to gradually change over time; trial tags just need to be resolved enough to distinguish the current trial from other recent trials; and position tags need only to distinguish between a small number of relevant serial positions. By contrast, stimuli can be highly confusable: storing stimuli in a way that distinguishes them from all other possible stimuli requires expensive coding for details. Putting all of this together, we hypothesize that the total ‘information load’ of maintaining the context representations in Figure 1 is no higher than (and potentially lower than) the information load that would be required to actively maintain the two stimuli. We should also point out that context can be actively represented even during periods of time when overall neural activity levels have returned to baseline; for example, any neural activity that drifts passively can serve as a temporally drifting context.

Another important caveat is that, while the particular configuration of context representations shown in Figure 1 is meant to be a reasonable example of how context-guided EM retrieval can solve the task, we certainly acknowledge that other configurations of context representations could exist that would give rise to the observed results, and that other factors could come into play. For example, although not shown in Figure 1, participants could strategically decide to place more of an attentional ‘weight’ on particular context dimensions at particular times in the

### Outstanding Questions

What are the roles of the HPC and MTL in dual retrocue studies? Can patients with damage to the HPC or surrounding MTL exhibit activity silent memory?

What effects do perturbations (e.g., the brief flashing of a task-irrelevant stimulus) have on HPC activity? Can reactivation of stimulus representations and/or signatures of neural plasticity in this region be observed following perturbations?

While we have focused on the roles of the HPC and MTL in supporting EM, we also emphasized that our definition of EM is functional (rapid acquisition of long-lasting, arbitrary associations). It remains an open question what other brain regions could contribute to the mechanisms of EM. For example, could other structures that exhibit these properties (e.g., cerebellum) be involved and/or is the PFC capable of long-term synaptic alterations that support EM trace retention?

Could gamma bursts reflect EM retrieval? Future neurophysiological studies could investigate the relationship between HPC and PFC gamma in the context of activity silent memory.

What are the neural substrates of the different kinds of context (trial tags, position codes, temporally drifting context) discussed here? Can we relate the similarity structure of the neural patterns in these regions to behavioral and neural measures of EM retrieval (e.g., PI, hippocampal activity)?

How are context representations learned and set up for a particular task? Different tasks will require different combinations of context representations. Future neural network models could provide insight into how our repertoire of context representations is established over development, how this repertoire is adapted to the task at hand, and whether it is possible to develop new types of context representations in the course of a single experiment.

### Box 2. Long-Term Memory (LTM) in Short-Term Memory Paradigms

While it is tempting to assume that performance on memory tasks with a short retention interval is necessarily supported by memory systems with short-lasting traces, our paper argues that this assumption may be false [40]; specifically, we argue that performance on WM tasks (which have a short retention interval) may also be supported by a memory system (EM) that stores lasting memory traces. Numerous studies have found that long-term memory systems (e.g., EM) can help or hinder performance in tasks that do not strictly require long-lasting memory traces [28,92–95], and memory reactivation has been shown to occur even in cases in which the memory item is no longer relevant for task performance [92,96,97]. These findings are consistent with the assumption that EM retrieval is automatically triggered when retrieval cues happen to match stored memory traces [50]. One classic example of how automatic, similarity-based retrieval can impair performance is the release from PI effect in the Brown–Peterson task [98,99]. If participants are given, over multiple trials, triplets from the same category to retain over a short delay period, performance gradually decreases over trials; however, if the category from which the triplets are drawn is changed, performance rapidly recovers [100]. These results show that EM can cause a build-up of PI that is released once the semantic features of the items change. In another example of EM hindering performance, [26] showed that, during simple (distraction free) maintenance of a set of target words, other words that were studied in the same context as the target words were automatically retrieved from EM, leading participants to confuse them with words that actually were studied and slowing reaction times. Under different circumstances, LTM can also help performance. For example, in [95] it was shown that even in a classic visual WM paradigm, the statistics of the stimuli from previous trials are retained, possibly in LTM. The authors propose a model in which such statistical learning sets a prior on what items are held in memory, helping to mitigate the degradation of WM representations. Also, although in [26] this was not demonstrated directly, the same context-based retrieval mechanism that led to intrusions of contextually related lures in their paradigm could also support the retrieval of contextually related targets if they happened to drop out of an actively maintained state. All of these findings highlight the importance of studying interactions between memory systems, even in (apparently) simple tasks with a short retention interval.

trial (e.g., upregulating position codes during the probes to ensure that the correct item is recalled). To adjudicate between these possibilities, it will be important to implement neural network models endowed with WM and EM [78–82] and apply them to specific WM tasks (like dual retrocue). These models can be used to develop hypotheses about how context representations are learned and/or selected under different task conditions. It will likewise be important to collect data to test the predictions of these models. Different configurations of context representations (trial codes, position codes, drift, etc.) lead to different predictions about the similarity of neural patterns across time points; these predictions could be tested by applying representational similarity analysis to the regions thought to maintain context and task-relevant representations (e.g., medial prefrontal cortex, orbitofrontal cortex, HPC; see [83]), and these representational similarity measures could then be related to behavioral measures of memory retrieval. Pursuing this approach should lead to a richer understanding of how WM and EM collaborate to support complex patterns of behavior on WM tasks [29] (Box 2).

### Acknowledgments

This work was supported by a grant from the John Templeton Foundation. We thank Matt Panichello and Taylor Webb for helpful discussion.

### Declaration of Interests

No interests are declared.

### References

1. Baddeley, A. (1992) Working memory. *Science* 255, 556–559
2. Constantinidis, C. *et al.* (2018) Persistent spiking activity underlies working memory. *J. Neurosci.* 38, 7020–7028
3. Fuster, J.M. and Alexander, G.E. (1971) Neuron activity related to short-term memory. *Science* 173, 652–654
4. Goldman-Rakic, P.S. (1995) Cellular basis of working memory. *Neuron* 14, 477–485
5. King, J.R. and Dehaene, S. (2014) Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn. Sci.* 18, 203–210
6. Zylberberg, J. and Strowbridge, B.W. (2017) Mechanisms of persistent activity in cortical circuits: possible neural substrates for working memory. *Annu. Rev. Neurosci.* 40, 603–627
7. Lewis-Peacock, J.A. *et al.* (2012) Neural evidence for a distinction between short-term memory and the focus of attention. *J. Cogn. Neurosci.* 24, 61–79
8. Rose, N.S. *et al.* (2016) Reactivation of latent working memories with transcranial magnetic stimulation. *Science* 354, 1136–1139
9. Stokes, M.G. *et al.* (2013) Dynamic coding for cognitive control in prefrontal cortex. *Neuron* 78, 364–375
10. Wolff, M.J. *et al.* (2017) Dynamic hidden states underlying working-memory-guided behavior. *Nat. Neurosci.* 20, 864–871
11. Kamiński, J. and Rutishauser, U. (2020) Between persistently active and activity-silent frameworks: novel vistas on the cellular basis of working memory. *Ann. N. Y. Acad. Sci.* 1464, 64–75



12. Masse, N.Y. *et al.* (2020) Reevaluating the role of persistent neural activity in short-term memory. *Trends Cogn. Sci.* 24, 242–258
13. Masse, N.Y. *et al.* (2019) Circuit mechanisms for the maintenance and manipulation of information in working memory. *Nat. Neurosci.* 22, 1159–1167
14. Miller, E.K. *et al.* (2018) Working memory 2.0. *Neuron* 100, 463–475
15. Mongillo, G. *et al.* (2008) Synaptic theory of working memory. *Science* 319, 1543–1546
16. Stokes, M.G. (2015) ‘Activity-silent’ working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn. Sci.* 19, 394–405
17. Sandberg, A. *et al.* (2003) A working memory model based on fast Hebbian learning. *Netw. Comput. Neural Syst.* 14, 789–802
18. Fiebig, F. *et al.* (2020) An indexing theory for working memory based on fast Hebbian plasticity. *eNeuro* 7 ENEURO.0374-19.2020
19. Watanabe, K. and Funahashi, S. (2014) Neural mechanisms of dual-task interference and cognitive capacity limitation in the prefrontal cortex. *Nat. Neurosci.* 17, 601–611
20. Cowan, N. (2001) The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* 24, 87–114
21. Oberauer, K. (2002) Access to information in working memory: exploring the focus of attention. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 411–421
22. Squire, L.R. and Zola-Morgan, S. (1991) The medial temporal lobe memory system. *Science* 253, 1380–1386
23. Ezzayat, Y. and Olson, I.R. (2008) The medial temporal lobe and visual working memory: comparisons across tasks, delays, and visual similarity. *Cogn. Affect. Behav. Neurosci.* 8, 32–40
24. Olson, I.R. *et al.* (2006) Visual working memory is impaired when the medial temporal lobe is damaged. *J. Cogn. Neurosci.* 18, 1087–1097
25. Goodrich, R.I. *et al.* (2019) Visual working memory impairments for single items following medial temporal lobe damage. *Neuropsychologia* 134, 107227
26. Hoskin, A.N. *et al.* (2019) Refresh my memory: episodic memory reinstatements intrude on working memory maintenance. *Cogn. Affect. Behav. Neurosci.* 19, 338–354
27. Baddeley, A. (2000) The episodic buffer: a new component of working memory? *Trends Cogn. Sci.* 4, 417–423
28. Foster, J.J. *et al.* (2019) Working memory as persistent neural activity. *PsyArXiv* Published online September 22, 2019. <https://doi.org/10.31234/osf.io/jh6e3>
29. Rose, N.S. (2020) The dynamic-processing model of working memory. *Curr. Dir. Psychol. Sci.* 29, 378–387
30. Wang, J.X. *et al.* (2014) Targeted enhancement of cortical-hippocampal brain networks and associative memory. *Science* 345, 1054–1057
31. Hermiller, M.S. *et al.* (2020) Evidence for immediate enhancement of hippocampal memory encoding by network-targeted theta-burst stimulation during concurrent fMRI. *J. Neurosci.* 40, 7155–7168
32. Lundqvist, M. *et al.* (2016) Gamma and beta bursts underlie working memory. *Neuron* 90, 152–164
33. Lundqvist, M. *et al.* (2011) Theta and gamma power increases and alpha/beta power decreases with memory load in an attractor network model. *J. Cogn. Neurosci.* 23, 3008–3020
34. Lundqvist, M. *et al.* (2018) Working memory: delay activity, yes! Persistent activity? Maybe not. *J. Neurosci.* 38, 7013–7019
35. Lundqvist, M. *et al.* (2018) Gamma and beta bursts during working memory readout suggest roles in its volitional control. *Nat. Commun.* 9, 394
36. Spellman, T. *et al.* (2015) Hippocampal-prefrontal input supports spatial encoding in working memory. *Nature* 522, 309–314
37. van Vugt, M.K. *et al.* (2010) Hippocampal gamma oscillations increase with memory load. *J. Neurosci.* 30, 2694–2699
38. Oberauer, K. *et al.* (2012) Modeling working memory: an interference model of complex span. *Psychon. Bull. Rev.* 19, 779–819
39. Wang, T.H. *et al.* (2019) More is less: increased processing of unwanted memories facilitates forgetting. *J. Neurosci.* 39, 3551–3560
40. Unsworth, N. and Engle, R.W. (2007) The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychol. Rev.* 114, 104–132
41. Jeneson, A. and Squire, L.R. (2012) Working memory, long-term memory, and medial temporal lobe function. *Learn. Mem.* 19, 15–25
42. McCabe, D.P. (2008) The role of covert retrieval in working memory span tasks: evidence from delayed recall tests. *J. Mem. Lang.* 58, 480–494
43. Rhodes, S. and Cowan, N. (2018) Attention in working memory: attention is needed but it yearns to be free. *Ann. N. Y. Acad. Sci.* 1424, 52–63
44. Cowan, N. (1999) An embedded-processes model of working memory. In *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, pp. 62–101, Cambridge University Press
45. Zaksas, D. and Pasternak, T. (2006) Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. *J. Neurosci.* 26, 11726–11742
46. Panichello, M.F. and Buschman, T.J. (2020) Selective control of working memory in prefrontal, parietal, and visual cortex. *bioRxiv* Published online April 9, 2020. <https://doi.org/10.1101/2020.04.07.030718>
47. Harrison, S.A. and Tong, F. (2009) Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458, 632–635
48. Serences, J.T. (2016) Neural mechanisms of information storage in visual short-term memory. *Vis. Res.* 128, 53–67
49. Nicoll, R.A. (2017) A brief history of long-term potentiation. *Neuron* 93, 281–290
50. Tulving, E. and Thomson, D.M. (1973) Encoding specificity and retrieval processes in episodic memory. *Psychol. Rev.* 80, 352–373
51. Homer, A.J. *et al.* (2015) Evidence for holistic episodic recollection via hippocampal pattern completion. *Nat. Commun.* 6, 7462
52. McClelland, J.L. *et al.* (1995) Why there are complementary learning systems in hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102, 419–457
53. Howard, M.W. and Kahana, M.J. (2002) A distributed representation of temporal context. *J. Math. Psychol.* 46, 269–299
54. Kahana, M.J. (2020) Computational models of memory search. *Annu. Rev. Psychol.* 71, 107–138
55. Manning, J.R. *et al.* (2014) Role of context in episodic memory. In *The Cognitive Neurosciences V*, MIT Press
56. Polyn, S.M. *et al.* (2009) A context maintenance and retrieval model of organizational processes in free recall. *Psychol. Rev.* 116, 129–156
57. Polyn, S.M. and Kahana, M.J. (2008) Memory search and the neural representation of context. *Trends Cogn. Sci.* 12, 24–30
58. Mensink, G.-J. and Raaijmakers, J.G. (1988) A model for interference and forgetting. *Psychol. Rev.* 95, 434–455
59. Cohen, J.D. *et al.* (1990) On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychol. Rev.* 97, 332–361
60. Desimone, R. and Duncan, J. (1995) Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222
61. Miller, E.K. and Cohen, J.D. (2001) An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202
62. Cohen, J.D. and O’Reilly, R.C. (1996) A preliminary theory of the interactions between prefrontal cortex and hippocampus that contribute to planning and prospective memory. In *Prospective Memory: Theory and Applications*, pp. 267–295, Lawrence Erlbaum Associates
63. Bulkin, D.A. *et al.* (2016) Placing memories in context: hippocampal representations promote retrieval of appropriate memories. *Hippocampus* 26, 958–971
64. Bulkin, D.A. *et al.* (2020) Hippocampal state transitions at the boundaries between trial epochs. *Hippocampus* 30, 582–595
65. Sun, C. *et al.* (2020) Hippocampal neurons represent events as transferable units of experience. *Nat. Neurosci.* 23, 651–663
66. Clewett, D. *et al.* (2019) Transcending time in the brain: how event memories are constructed from experience. *Hippocampus* 29, 162–183

67. DuBrow, S. *et al.* (2017) Does mental context drift or shift? *Curr. Opin. Behav. Sci.* 17, 141–146
68. Steemers, B. *et al.* (2016) Hippocampal attractor dynamics predict memory-based decision making. *Curr. Biol.* 26, 1750–1757
69. Wen, T. *et al.* (2020) Hierarchical representation of multi-step tasks in multiple-demand and default mode networks. *J. Neurosci.* 40, 7724–7738
70. Liu, Y. *et al.* (2019) Human replay spontaneously reorganizes experience. *Cell* 178, 640–652.e14
71. Morton, N.W. *et al.* (2020) Representations of common event structure in medial temporal lobe and frontoparietal cortex support efficient inference. *Proc. Natl. Acad. Sci. U. S. A.* 117, 29338–29345
72. MacDonald, C.J. *et al.* (2011) Hippocampal “time cells” bridge the gap in memory for discontinuous events. *Neuron* 71, 737–749
73. Taxisis, J. *et al.* (2020) Differential emergence and stability of sensory and temporal representations in context-specific hippocampal sequences. *Neuron* 108, 984–998.e9
74. Mau, W. *et al.* (2018) The same hippocampal CA1 population simultaneously codes temporal information over multiple timescales. *Curr. Biol.* 28, 1499–1508.e4
75. Tiganj, Z. *et al.* (2018) Compressed timeline of recent experience in monkey lateral prefrontal cortex. *J. Cogn. Neurosci.* 30, 935–950
76. Wang, Y. *et al.* (2006) Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nat. Neurosci.* 9, 534–542
77. Chaudhuri, R. and Fiete, I. (2016) Computational principles of memory. *Nat. Neurosci.* 19, 394–403
78. Beukers, A. *et al.* (2019) Working with episodic memory: the *n*-back task. In *2019 Conference on Cognitive Computational Neuroscience, CCN*
79. Graves, A. *et al.* (2014) Neural Turing machines. *arXiv* Published online October 20, 2014. <http://arxiv.org/abs/1410.5401>
80. Graves, A. *et al.* (2016) Hybrid computing using a neural network with dynamic external memory. *Nature* 538, 471–476
81. Ritter, S. *et al.* (2018) Been there, done that: meta-learning with episodic recall. *arXiv* Published online May 24, 2020. <http://arxiv.org/abs/1805.09692>
82. Ritter, S. *et al.* (2018) Episodic control as meta-reinforcement learning. *arXiv* Published online July 3, 2018. <https://doi.org/10.1101/360537>
83. Schuck, N.W. and Niv, Y. (2019) Sequential replay of nonspatial task states in the human hippocampus. *Science* 364, eaaw5181
84. Estes, W.K. (1955) Statistical theory of spontaneous recovery and regression. *Psychol. Rev.* 62, 145–154
85. Howard, M.W. and Eichenbaum, H. (2013) The hippocampus, time, and memory across scales. *J. Exp. Psychol. Gen.* 142, 1211–1230
86. Folkerts, S. *et al.* (2018) Human episodic memory retrieval is accompanied by a neural contiguity effect. *J. Neurosci.* 38, 4200–4211
87. Manning, J.R. *et al.* (2011) Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. *Proc. Natl. Acad. Sci. U. S. A.* 108, 12893–12897
88. Shin, Y.S. and DuBrow, S. (2020) Structuring memory through inference-based event segmentation. *Top. Cogn. Sci.* Published online May 27, 2020. <https://doi.org/10.1111/tops.12505>
89. Sahakyan, L. and Kelley, C.M. (2002) A contextual change account of the directed forgetting effect. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 1064–1072
90. Manning, J.R. *et al.* (2016) A neural signature of contextually mediated intentional forgetting. *Psychon. Bull. Rev.* 23, 1534–1542
91. Behrens, T.E.J. *et al.* (2018) What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron* 100, 490–509
92. Barbosa, J. *et al.* (2020) Interplay between persistent activity and activity-silent dynamics in the prefrontal cortex underlies serial biases in working memory. *Nat. Neurosci.* 23, 1016–1024
93. Mizrak, E. and Oberauer, K. (2020) Working memory recruits long-term memory when it is beneficial: evidence from the Hebb effect. *PsyArXiv* Published online May 11, 2020. <https://doi.org/10.31234/osf.io/ryqj9>
94. Ngiam, W.X.Q. *et al.* (2019) “Memory compression” effects in visual working memory are contingent on explicit long-term memory. *J. Exp. Psychol. Gen.* 148, 1373–1385
95. Panichello, M.F. *et al.* (2019) Error-correcting dynamics in visual working memory. *Nat. Commun.* 10, 3366
96. van Loon, A.M. *et al.* (2018) Current and future goals are represented in opposite patterns in object-selective cortex. *eLife* 7, e38677
97. Bae, G. and Luck, S.J. (2019) Reactivation of previous experiences in a working memory task. *Psychol. Sci.* 30, 587–595
98. Brown, J. (1958) Some tests of the decay theory of immediate memory. *Q. J. Exp. Psychol.* 10, 12–21
99. Peterson, L.R. and Peterson, M.J. (1959) Short-term retention of individual verbal items. *J. Exp. Psychol.* 58, 193–198
100. Wickens, D.D. (1970) Encoding categories of words: an empirical approach to meaning. *Psychol. Rev.* 77, 1–15