

Review

Reevaluating the Role of Persistent Neural Activity in Short-Term Memory

Nicolas Y. Masse ^{1,3,4,*} Matthew C. Rosen,^{1,4} and David J. Freedman^{1,2,*}

A traditional view of short-term working memory (STM) is that task-relevant information is maintained ‘online’ in persistent spiking activity. However, recent experimental and modeling studies have begun to question this long-held belief. In this review, we discuss new evidence demonstrating that information can be ‘silently’ maintained via short-term synaptic plasticity (STSP) without the need for persistent activity. We discuss how the neural mechanisms underlying STM are inextricably linked with the cognitive demands of the task, such that the passive maintenance and the active manipulation of information are subserved differently in the brain. Together, these recent findings point towards a more nuanced view of STM in which multiple substrates work in concert to support our ability to temporarily maintain and manipulate information.

Persistent but Not Consistent: Delay-Period Spiking and STM

In our day-to-day lives, the information we need to make informed decisions and actions rarely arrives all at once (e.g., holding a conversation, reading this review). Rather, it often arrives piecemeal, requiring us to link together temporally disjointed bits of information to obtain a more complete picture of our environment. Doing so requires not only the temporary maintenance of information in memory but also manipulation and transformation of the information into different representations or actions. These cognitive functions are crucial for higher-order intelligence generally [1–3] and thus understanding the neural mechanisms that subserve them has been a longstanding goal in neuroscience.

According to the textbook view, information in STM is maintained in stimulus-selective persistent activity, commonly observed in the parietal and frontal cortices (See Box 1) [4–7]. However, there is a growing realization that this traditional picture is incomplete [8,9], as several electrophysiology studies have demonstrated that the strength of persistent activity can be highly variable, weak, or even entirely absent during tasks that require STM [10–16]. Furthermore, human functional imaging studies have shown that persistent activity can disappear if the memoranda are not immediately behaviorally relevant but can ‘reawaken’ as soon as the task requires [17,18]. These findings raise two fundamental questions: (i) why does persistent activity vary between tasks; and (ii) for those tasks with weak or nonexistent persistent activity, how is information maintained?

Thanks to recent conceptual and technical advances, answers to both questions have rapidly begun to materialize. Several groups have built on theoretical work [19] to provide evidence that information can be temporarily maintained at the synaptic level via STSP without the need for persistent spiking activity [20–22]. These experiments have also begun to elucidate how the presence of persistent activity depends on the task’s cognitive demands. Mounting evidence supports the growing consensus that the neural mechanisms underlying the passive maintenance of information differ from those involved in the active manipulation of remembered information [23]. To make our discussion of this distinction clear, we adopt a common nomenclature: STM denotes the broad set of cognitive processes that require the temporary maintenance of

Highlights

It has been commonly believed that information in short-term memory (STM) is maintained in persistent delay-period spiking activity.

Recent experiments have begun to question this assumption, as the strength of persistent activity appears greater for tasks that require active manipulation of the memoranda, as opposed to tasks that require only passive maintenance.

New experiments have revealed that information in STM can be maintained in neural ‘hidden’ states, such as short-term synaptic plasticity.

Machine-learning-based recurrent neural networks have been successfully trained to solve a diversity of working memory tasks and can be leveraged to understand putative neural substrates of STM.

¹Department of Neurobiology, The University of Chicago, Chicago, IL, USA

²Grossman Institute for Neuroscience, Quantitative Biology and Human Behavior, The University of Chicago, Chicago, IL, USA

³Current address: Kebotix, Inc., Cambridge, MA, USA

⁴These authors contributed equally

*Correspondence: nicolas.masse@gmail.com (N.Y. Masse) and dfreedman@uchicago.edu (D.J. Freedman).



Box 1. Persistent Activity: Where and for What

It is currently debated which cortical areas maintain memoranda-specific information in STM. While single-unit recording studies in nonhuman primates clearly demonstrate the existence of persistent activity in the PFC [4–7,12] and parietal cortex [5,14], similar studies failed to reveal its presence in more posterior areas, such as the visual cortex [16,89–91]. This led to the early belief that STM is solely based in the frontoparietal network. However, more recent fMRI [56,92] and EEG [93] studies have demonstrated that stimulus information can be decoded from posterior areas during the STM delay period, raising questions about where in the brain STM information is maintained.

The discrepancy between these two results can be explained by the fact that fMRI and EEG signals more closely reflect synaptic potentials, as opposed to spiking activity [94]. Specifically, recent studies have shown that STM can modulate LFP oscillations in posterior areas [95], which does not necessarily change the level of spiking activity but would drive EEG and fMRI BOLD signal activity. Furthermore, spiking activity in frontal areas drives these posterior LFPs [96,97], which act to coordinate spike timing in posterior areas [98]. This suggests that, when high-fidelity sensory information needs to be held in STM, frontoparietal delay activity might be more involved in supporting the maintenance of information in STM, whereas the information resides in posterior areas [99]. However, if coarse-grained or categorical information needs to be maintained, studies suggest that frontoparietal networks can maintain stimulus information [11,37].

Identifying the cortical areas involved in the various cognitive aspects of STM remains a work in progress. However, this body of work highlights the fact that STM function is highly task dependent and that distinguishing between persistent activity that is associated with information maintenance and persistent activity associated with other cognitive aspects of STM (manipulation, attention, etc.), remains a highly nontrivial problem.

information. When referring specifically to those STM processes that also demand active manipulation of memoranda, we use the term working memory (WM) [24–26].

Here, we review recent studies that have illuminated the roles of both persistent spiking activity and STSP in STM and WM. We discuss how these representations depend on the specific cognitive demands of different tasks and how experiments have begun to differentiate the neural mechanisms underlying the maintenance and manipulation of information. After defining the particular challenges that this emerging picture poses, we conclude by highlighting a burgeoning class of ‘model organisms’ that appears uniquely poised to help: machine-learning-based recurrent neural networks (RNNs).

Stable Persistent Activity

Much of our knowledge of the neural mechanisms underlying STM and WM derives from experiments requiring a delayed response to a stimulus, like the now-classic memory-guided saccade task (Figure 1A). Here, the subject must remember the spatial location of a visual target across a ~1-s delay period, before making a saccadic eye movement towards the remembered location after the fixation cue is extinguished. Figure 1B shows the response of an example neuron recorded from the parietal cortex that persistently and selectively fires throughout the delay period for visual targets flashed in the upwards location. Such stimulus-selective persistent firing through the delay period of STM and WM tasks has been widely observed over the past several decades [4–7]. Hopfield reinforced this experimental evidence in the early 1980s by reformulating STM in a dynamical systems perspective (Box 2) [27]. Specifically, he demonstrated that neural circuits can form point attractors in state space to store discrete memories. Persistent activity arises as a natural byproduct of this process.

However, direct evidence that neural circuits form such discrete attractors has remained sparse. A recent study [28] addressed this gap experimentally using both intracellular and extracellular recordings from the anterior lateral motor cortex (ALM) of mice performing a delayed-response task using either tactile or auditory stimuli. Despite previous work [29,30] showing that the ALM supports persistent activity, the mechanism underlying the emergence of the discrete attractors remained unclear. The authors first used intracellular recordings to rule out the possibility that

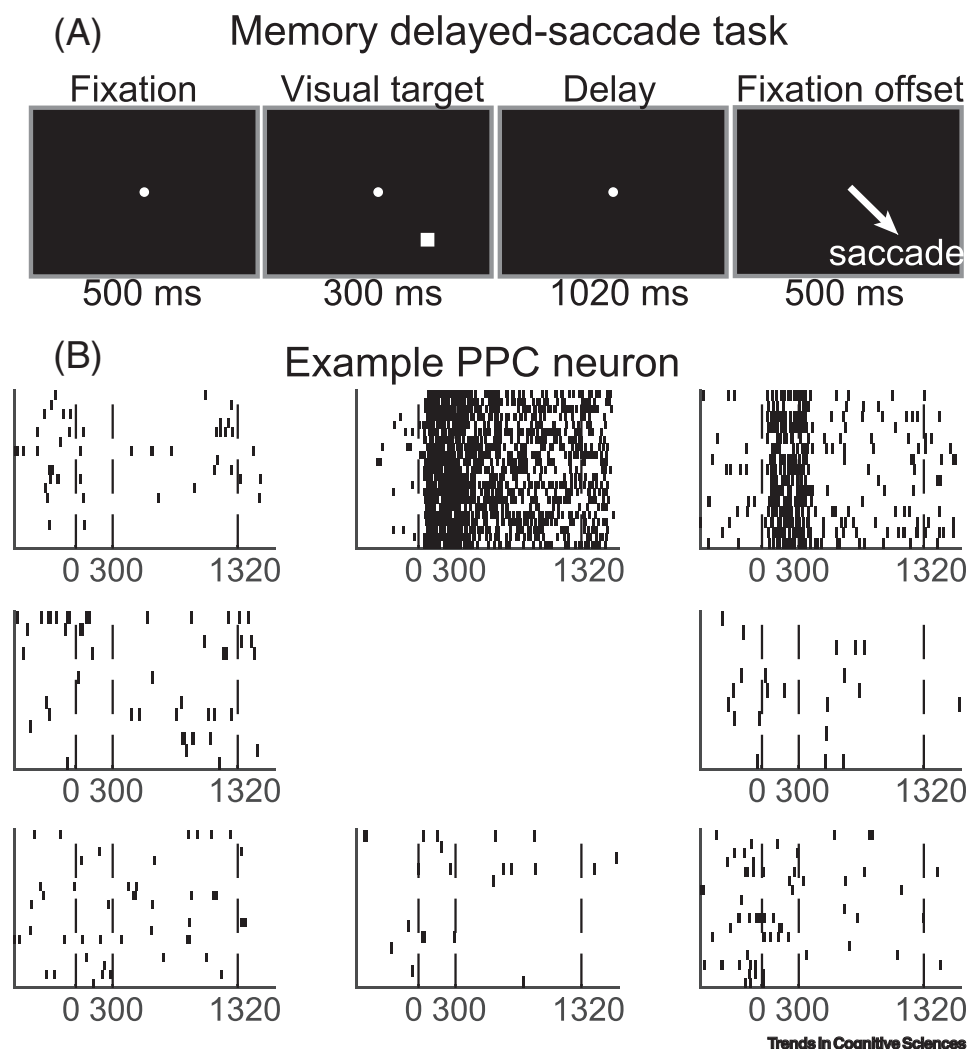


Figure 1. Persistent Activity. Example neuron showing clear delay-period persistent activity. (A) This neuron was recorded during the widely used memory-guided saccade task, in which the subject had to remember the spatial location of a visual target across a ~1-s delay period before making a saccadic eye movement towards the remembered location after the fixation cue is turned off. (B) Spike raster (each row is a trial, each vertical line is an action potential) from an example neuron from the posterior parietal cortex showing persistent spiking activity throughout the delay period for visual targets flashed in the upwards location. Neuron taken from experiment in [15].

cell-autonomous properties (e.g., long membrane time constant; reviewed in [31]) could explain the observed persistent activity, implicating network connectivity in the establishment of persistence. Furthermore, extracellular recordings from ALM neurons showed that population activity evolves towards one of two distinct endpoints during the delay, each of which corresponded with a distinct behavioral response. Crucially, after optogenetic perturbations, neural activity either (i) returned to the same endpoint it originated from, and the animal performed the same behavioral response as expected, or (ii) settled towards the other endpoint, and the animal switched behavioral responses.

One important caveat is that the animal could solve the task by maintaining a representation of the sensory stimulus throughout the delay period or by rapidly transforming the sensory

Box 2. A State Space Perspective of STM

What is the best way to conceptualize how neural activity, and the neural circuits that support it, allows the maintenance and manipulation of information in STM? Here, we follow the lead of others [8,83–87] and adopt the language of neural state spaces, in which the neural activity of a population of N neurons can be described as points in N -dimensional Euclidean space. The coordinate in dimension i at any timepoint is given by the activity of neuron i at that timepoint. Effective connectivity between neurons – how effectively spiking by one neuron drives spiking by another – determines the trajectories that population activity traces out in this N -dimensional space. Thus, the wiring circuit determines which paths within state space are possible [88].

Consider a simple example network of neurons: three excitatory cells and two inhibitory cells, shown in Figure 1A. Neurons E1 and E2 mutually excite each other, as do neurons E2 and E3; when E1 is active, it also excites neuron I1, which inhibits E3 (Stimulus #1, Figure 1B). A similar pattern holds when E3 is active. This is an example of a competitive circuit, where E1 and E3 cannot be co-active. The hypothetical operation of this circuit is illustrated in the rasters shown in Figure 1B. Because E2 is always active, two dimensions are sufficient to capture the structure of the state space that this circuit describes (Figure 1C). The dynamics of this competitive circuit form two discrete point attractors, or stable points in state space towards which the circuit's connectivity drives activity.

To better understand how neural activity evolves through this space, we can assign every point in the landscape to an energy value. In the same way that gravity pulls a ball down the slope of a valley to its base, trajectories in state space tend to move in directions that minimize energy. Figure 1D schematizes the energy landscape associated with the neural circuit described in Figure 1A, in which yellow colors represent high-energy states and blue colors represent low-energy states. Changing the connectivity of a circuit changes the topography of its energy landscape, which in turn changes how neural trajectories progress through state space.

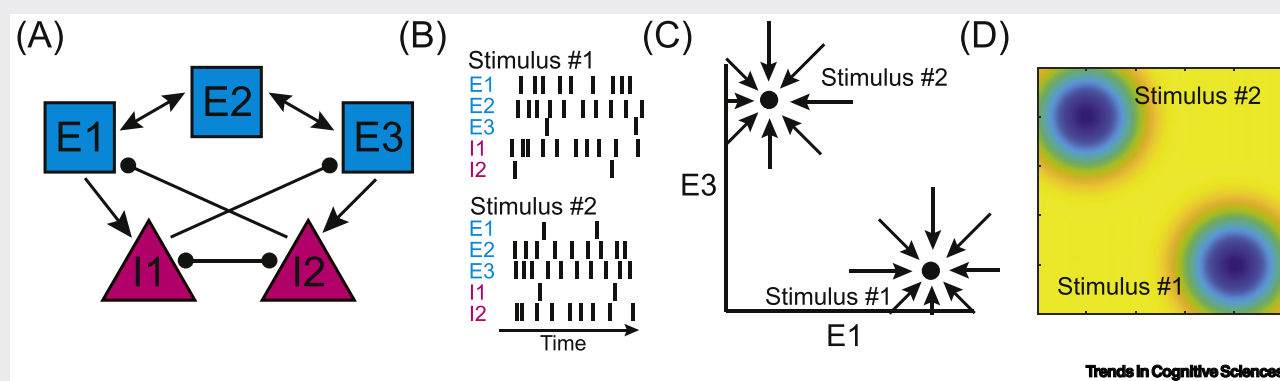


Figure 1. State-Space Schematic. Schematic illustrating the association between stimulus-specific persistent activity, synaptic connectivity, and the energy landscape. (A) Hypothetical network of three excitatory (E) and two inhibitory (I) neurons. Arrowheads and circles indicate excitatory and inhibitory connections, respectively. (B) Hypothetical raster associated with the network shown in (A). (C) The fixed-point attractors associated with stimuli #1 and #2 can be observed by projecting the state space onto the 2D subspace spanned by the neural activity of E1 (x-axis) and the neural activity of E3 (y-axis). (D) The energy landscape showing the two attractors in state space. Yellow is associated with higher-energy states and blue with lower-energy states. Patterns of neural activity in state space will tend towards lower-energy states.

stimulus into a motor plan, which it maintains through the delay period. This muddies the interpretation of the precise role of persistent activity: it is unclear whether it encodes sensory information, the motor plan to be executed, or possibly some combination of the two. Whether persistent activity relates to information maintenance or to other cognitive functions is a theme we continue to explore in subsequent sections.

Dynamic Persistent Activity

The results described above provide evidence that networks can instantiate distinct attractors in state space that correspond with stimulus-specific behavioral responses. However, many other studies of STM find delay-period activity with temporal dynamics that deviate from orbits around fixed points [10,32,33]. For example, in a study that dissociated the stimulus location from the upcoming saccade location in a memory-guided saccade-style task, activity in the frontal cortex initially encoded the stimulus location (retrospective code) before its representation shifted towards encoding the planned saccade target (prospective code) later in the delay [34].

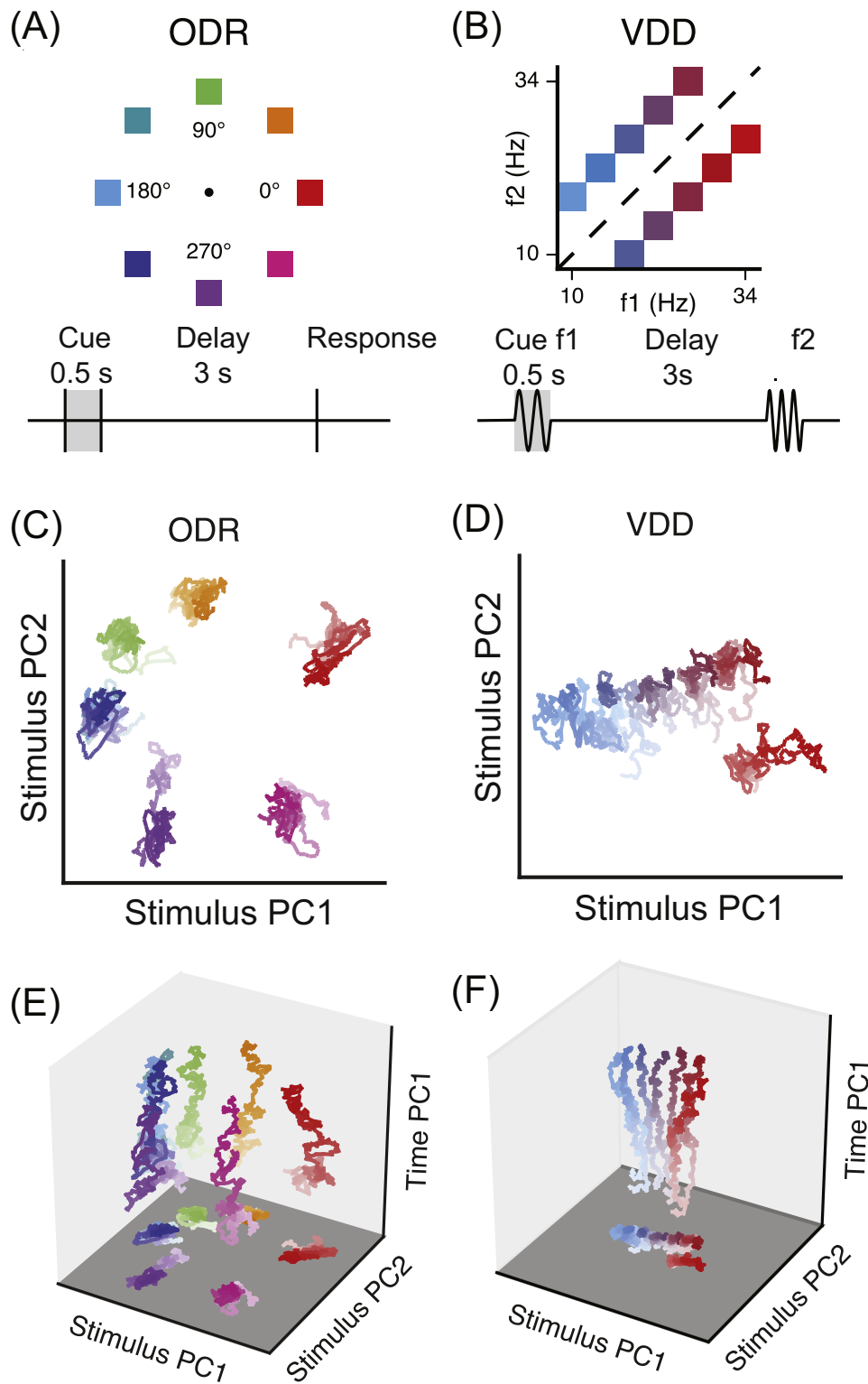
How can the discrete attractor story be consistent with neural trajectories that are not localized in neural state space? Would this kind of dynamics pose problems for downstream readout of memory contents? One possible solution is that network attractors are low-dimensional subspaces rather than single points. Thus, delay-period activity can adopt rich, varied dynamics as it evolves along a low-dimensional subspace attractor. This solution was originally proposed by modeling work showing that subspace attractors can allow the maintenance of information in STM while explaining the patterns of dynamic neural activity during the delay period [11].

A recent study demonstrates the existence of these subspace attractors in monkeys performing two canonical STM tasks: an oculomotor delayed-response task (ODR) (Figure 2A), in which the monkey remembered the location of a visual stimulus before making a saccadic eye movement towards the remembered location (same task as shown in Figure 1); and a vibrotactile delayed discrimination task (VDD) (Figure 2B), in which the animal reports whether the vibrational frequency of a test stimulus is greater or less than the vibrational frequency of a previously presented sample stimulus [35]. The authors found, for both tasks, that neural activity in the prefrontal cortex (PFC) evolved along two orthogonal subspaces. In one subspace (the 'mnemonic' subspace), neural trajectories encoded the stimulus identity but remained mostly invariant, or stable, across time (Figure 2C,D). In the orthogonal subspace (the 'dynamic' subspace), trajectories evolved across time but remained invariant across stimulus conditions (Figure 2E,F). Thus, dynamic population activity during the memory delay period is consistent with the existence of subspace attractors in which stimulus information can be stably encoded across time in one subspace yet vary dynamically in another subspace. These distinct subspaces confer distinct theoretical advantages on STM. The existence of a mnemonic subspace that supports stable decoding of stimuli makes the job of downstream circuits easier: one set of weights is sufficient to extract the encoded information across all timepoints. The existence of a separate dynamic subspace, however, allows the circuit retaining information for STM more flexibility in subserving other, concurrent cognitive functions.

STM without Persistent Activity

While many studies have documented the existence of persistent activity during STM-based tasks, recent experimental work has demonstrated that, in many cases, the strength of persistent activity in the frontoparietal network can vary markedly and is often extremely weak, transient, or even absent [10–16]. In a delayed image-sequence-matching task, for example, stimulus encoding by spiking activity in the PFC is extremely weak during the delay period, rising only (i) during sample image presentation and (ii) during the test period, when the animal uses its memory to generate a motor response [36]. These results mirror those of human functional imaging studies that have shown that memoranda that are not immediately behaviorally relevant can no longer be decoded from the BOLD signal but can reemerge if they become relevant later in the task [17,18].

A recent study from our group demonstrated directly how persistent activity appears during some tasks but not others. This work showed that delay-period encoding in the lateral intraparietal (LIP) area was weak during a delayed motion-direction-matching task [14,15], but after the same monkeys underwent extensive categorization training using the same stimuli, delay-period encoding became highly robust [14]. Thus, persistent activity emerged only after the subject was required to transform the stimulus into a categorical representation. Both matching and categorization tasks used identical visual stimuli, timings of task events, and motor responses. Furthermore, the neuronal recordings were conducted from the same regions of the LIP in the two tasks, strongly suggesting that the emergence of persistent activity was primarily due to the change in the task demands.



Trends in Cognitive Sciences

(See figure legend at the bottom of the next page.)

Interestingly, persistent activity emerges upstream of the LIP, in the middle superior temporal (MST) area, during a similar (but not identical) delayed motion-direction-matching task [37]. This could reflect an additional demand of the task used in that study: the sample and test stimuli were presented in different retinotopic locations, forcing the subject to translate information from the sample location to the location of the test stimulus. By contrast, in our LIP studies, sample and test stimuli were always shown at the same retinotopic location and thus both activated the same pools of sensory neurons with receptive fields that overlapped with the stimuli. In theory, this overlap would enable a purely synaptic form of STM to maintain information in our LIP studies. Because the same feedforward pathways process both sample and test stimuli, potentiation and/or depression of those synapses conveying information about the sample could be sufficient to sculpt the processing of the test stimulus to generate the correct behavioral response.

In all of these cases, the strength and time course of persistent activity depend strongly on the sensory, motor, or cognitive demands of the task. This begs two important questions: why does persistent activity emerge in some cases but not in others; and what does this say about persistent activity's functional role in STM?

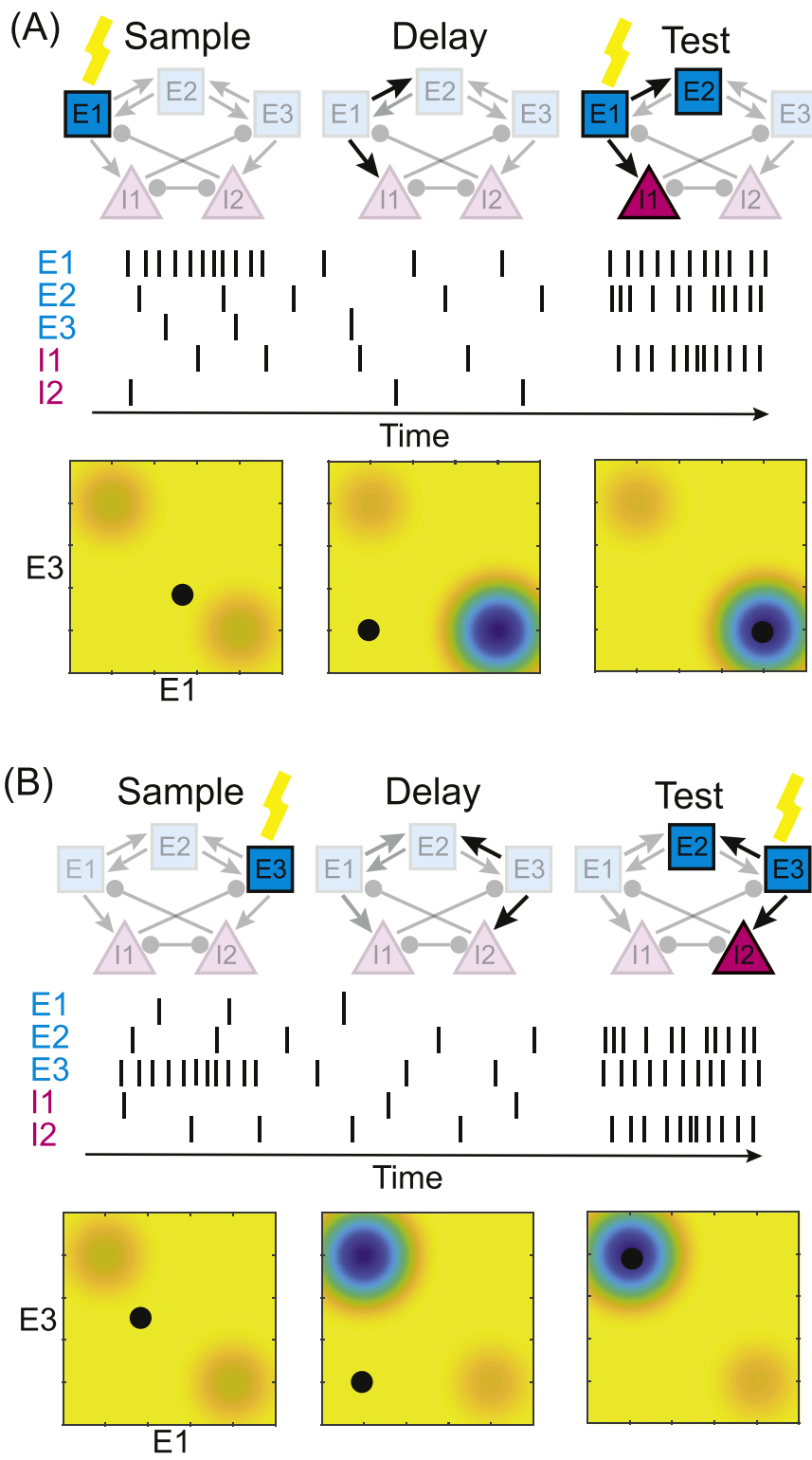
Activity-Silent Maintenance of Information in STM

The converging evidence described above suggests that persistent delay encoding might fail to tell the full story of STM. While spiking activity plays a central role, other neuronal substrates may have historically been overlooked because of limitations in how we measure neural activity. Most of our neuronal recording approaches measure signals related to the spiking input into an area (e.g., EEG, MEG, fMRI) or neurons' spiking output, either directly (e.g., electrophysiology) or by proxy (e.g., two-photon calcium imaging). What if other physical substrates in the brain could also temporarily maintain information?

This was the focus of an important modeling study, which showed in theory that STSP can maintain information without delay activity [19]. Synaptic efficacies can be highly dynamic and are either facilitated or depressed depending on presynaptic activity. Thus, stimuli to be maintained in STM will elicit different patterns of spiking activity, which will facilitate or depress different patterns of synapses, with changes in synaptic efficacy lasting hundreds to thousands of milliseconds.

Figure 3A,B illustrates this idea with a toy example. Consider a simple network of three excitatory and two inhibitory neurons (similar to Figure 1A in Box 2), initialized with weak connections as indicated by the gray arrows. Suppose a brief sample stimulus excites neuron E1 in Figure 3A and a different sample stimulus excites neuron E3 in Figure 3B. The initially weak connections corresponding to attractors in neural state space are equally weak (indicated by the energy landscapes in the middle panels that are almost uniformly yellow) and thus the neural response of E1 (Figure 3A) or E3 (Figure 3B) does not excite neighboring neurons, preventing the emergence

Figure 2. Stimulus Encoding via Subspace Attractors. The existence of a subspace attractor in two short-term memory (STM)-based tasks. (A) Oculomotor delayed-response (ODR) task (similar to the memory delayed-saccade task in Figure 1A). Subject has to remember the location of a visual target (shown by the colored squares) across a 3-s delay period. (B) In the vibrotactile delayed-discrimination (VDD) task, the subject was presented with two vibrotactile stimuli of different frequencies (f_1 and f_2) separated by a 3-s delay period. The subject had to indicate whether the test frequency (f_2) was greater than the sample frequency (f_1). (C) Population trajectories during the delay period projected onto the first two principal components (PC1 on the x-axis, PC2 on the y-axis) for the ODR task. Each color represents one stimulus condition. (D) Same as (C), except for the VDD task. (E) Neural trajectories projected onto a 3D subspace for the ODR task. x- and y-axes are the same as in (C), while the z-axis captures time-related variance in the trajectories. (F) Same as (E), except for the VDD task. Adapted from [35].



(See figure legend at the bottom of the next page.)

of persistent activity through the delay period. However, short-term synaptic facilitation would allow the sample stimulus to temporarily facilitate connections from E1 (Figure 3A) or from E3 (Figure 3B), with facilitation lasting throughout the delay period. This forms a memory trace in the effective synaptic connections (black arrows) and, correspondingly, in the energy landscapes, allowing the network to recognize a matching test stimulus.

Experiments have only recently begun to offer evidence supporting this idea that STSP can allow for information maintenance. In one of these, human participants had to remember the spatial location of two visual stimuli [22]. Immediate cueing of the participant about which stimulus to attend resulted in accurate decoding of that stimulus's spatial location from the participant's BOLD signal during the delay period. If neither stimulus was cued, the decoding accuracy for both stimuli decayed towards baseline levels. However, retrospective cueing of one stimulus as relevant for the response resulted in a significant increase in the decoding accuracy for the newly attended stimulus. This suggests that the decrease in decoding accuracy before the retro-cue was not because stimulus information was completely lost, but rather that it was stored in a latent state, hidden from the BOLD measurement. When the retro-cue was presented, this information was then reactivated into a measurable signal.

While this study indirectly shows that non-attended information can be stored in a hidden state, stronger evidence would entail direct measurement of the hidden states' contents. This was the goal of two recent studies with roughly similar behavioral experiments as the one above. In the first of these [21], human subjects were simultaneously presented with two oriented sample patterns and instructed to remember their orientations (Figure 4A). The subjects were instructed before the start of the trial which of the two sample stimuli was relevant for the first probe and which for a subsequent probe. Thus, the subjects needed to remember both sample stimuli during the first delay, but had to remember only the sample stimulus relevant for probe #2 during the second delay. During each of the two probes, the subjects had to indicate whether the probe was rotated clockwise or counterclockwise from the relevant sample stimulus.

The novel element of this task was that the contents of STM were 'pinged' by presenting a high-contrast, task-irrelevant visual impulse during the first and second delay periods. The idea is that if activity-silent hidden states are encoding the visual stimuli in WM, the hidden states will differ after different stimulus presentations. It follows that a subsequent stimulus would elicit different neural responses depending on the hidden contents of STM. This is what the authors observed. While the decoding accuracy of both visual orientations (measured using EEG activity) slowly decayed after stimulus offset (Figure 4B), pinging the subject partially rescued the decoding accuracies of both stimuli during the first delay (Figure 4C) and partially rescued the decoding accuracy of second stimulus to be tested (Figure 4D). Thus, the authors found direct support for the notion that information in STM can be maintained in hidden neural states whose contents can be revealed by pinging the network with task-irrelevant stimuli.

Figure 3. Model of Short-Term Synaptic Plasticity (STSP)'s Role in Short-Term Memory (STM). Schematic illustrating how STSP can maintain stimulus-specific information in STM. (A) Top panel: Similar hypothetical network as in Figure 1A in Box 2, except that gray arrows indicate that synaptic connections are initially weak. A sample stimulus excites neuron E1, which facilitates its synaptic connections onto E2 and I1 (indicated by the black arrows). These facilitated synapses create an attractor (shown by the energy landscape in the bottom panels) that persists throughout the delay period. This attractor allows the network to recognize a matching test stimulus. (B) Same as (A), except for a different sample stimulus that excites neuron E3. This creates an attractor in a different part of state space, showing that STSP can create stimulus-specific attractors that persist throughout the delay period even without persistent delay activity.

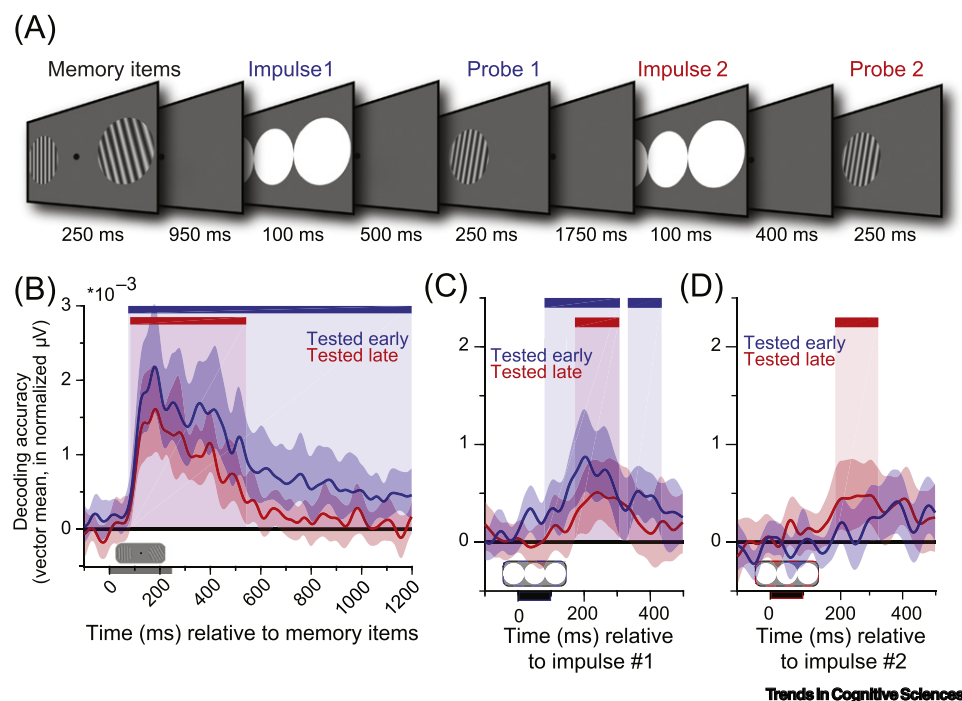


Figure 4. 'Hidden State' Stimulus Encoding. Decoding information maintained in hidden neural states during short-term memory (STM). (A) Subjects had to remember two items in STM, which were to be tested in a known order during probes 1 and 2. Impulses 1 and 2 were behaviorally irrelevant stimuli meant to 'ping' the hidden contents of STM. (B) Decoding accuracy of the memory item to be tested during probe #1 (blue) and item to be tested during probe #2 (red). (C) Decoding accuracy of the two memory items after impulse #1. Decoding accuracy for both items increased above chance after the impulse. (D) Decoding accuracy after the second impulse. Decoding accuracy for the remaining item to be tested increased above chance. Adapted from [21].

This approach was also used in another, technically impressive study [20]. Here, human subjects had to remember two different types of items (from groups of faces, words, or motion patterns) and the subjects were cued on which item was to become behaviorally relevant. The twist was that the authors used fMRI analysis to identify the cortical area encoding each type of stimulus and then pinged that area with a pulse of transcranial magnetic stimulation (TMS). As above, the TMS ping also partially rescued the decoding accuracy, further supporting the notion that hidden neural states can maintain information in STM.

It is worth restating that STM, and in particular WM, can involve both the maintenance and the manipulation of information. While the experiments described above support the hypothesis that activity-silent hidden states can support information maintenance, it was unknown whether they can also support activity-silent manipulation.

This idea was tested by a study [38] using a task in which human subjects were presented with a brief (17 ms) visual target location, which was near perceptual threshold and thus barely perceptually visible to the participants. The subjects were then cued to report (i) the exact target location, (ii) a location 120° clockwise from the target, or (iii) a location 120° counterclockwise from the target. At the end of the trial, the subjects rated the subjective visibility of the stimulus. During the trial, the position of the target flash was decoded from MEG activity. On trials in which the targets were perceptually visible, target decoding accuracy persisted throughout the trial, suggesting that target information was maintained by persistent delay activity. In contrast, the

decoding accuracy of targets that were not perceptually visible was weaker and quickly decayed to chance levels. Importantly, before the cue appeared indicating which manipulation to perform, one could decode the ‘pre-rotation location’ (the estimated target location consistent with the subject’s response) from the MEG signal, indicating that a representation of the target location, which could be decoded only at chance level for most of the trial, was reinstated back into delay activity. Thus, in contrast to the studies described above demonstrating passive maintenance without delay-period spiking, this study shows that active manipulation of information in WM requires delay activity. This result also reinforces the view that delay-period activity can be highly dependent on the task demands.

These fMRI, EEG, and MEG studies all share an important limitation: they measure the aggregate spiking activity pooled across millions of neurons and thus might lack the resolution to discern whether unattended information is maintained in a purely activity-silent manner. A recent human neuroimaging study [39] shows that unattended memoranda can still be weakly, but significantly, decoded from the BOLD signal from parietal and frontal cortices. One possibility is that information is not necessarily maintained in either persistent activity or STSP but in some combination of the two substrates. Models have suggested that low levels of persistent activity can help to ‘refresh’ information maintained in STSP [19]. Regardless, the development of better recording technologies to measure STSP *in vivo* will be required to definitively answer these questions.

Other Substrates Supporting STM and WM

STSP is one of potentially many substrates with the capacity to temporarily maintain information in memory. In theory, any biochemical or physiological process whose current state reflects its recent history, and which can be used by the organism to guide behavior [40], could fill this role. For example, modeling work has demonstrated that short-term Hebbian plasticity, in which coordinated pre- and post-synaptic firing temporarily increases synaptic efficacy, can allow stimulus-specific mnemonic encoding [41,42]. In another example, experimental and modeling studies have demonstrated that NMDA receptors can also support STM with their slow channel kinetics and voltage-gated channels [43–46]. Last, diffuse projections of neuromodulators like dopamine and acetylcholine from midbrain areas [47–51] can also influence STM by mediating the balance between network stability (e.g., invariance to new stimuli) and flexibility (e.g., the integration of new stimuli) [50,52,53].

While these substrates discussed above can allow information maintenance, other substrates and mechanisms might be more related to how information is coordinated. STM is highly distributed across the brain [54,55]: neural correlates of STM have been observed in the frontal [4–7,12], parietal [5,14], occipital [56,57], and temporal cortices [58], the medial temporal lobe [59,60], and the thalamus [61,62]. This raises an important question: if information maintained in STM needs to be manipulated, how does the brain properly coordinate and route the flow of information across multiple cortical and subcortical areas? One widely proposed solution is that coherent oscillation between cortical areas can regulate and coordinate the flow of information. While we do not have the space here to discuss recent studies detailing how oscillatory dynamics support STM, several publications [63,64] provide an excellent review of this topic.

This list of mechanisms of STM and WM is hardly exhaustive [40]. At this time, we are unsure whether information in STM is distributed across a large class of different substrates that all work in concert or whether it is primarily concentrated in a few. The data suggest not only that persistent spiking is one of several substrates of STM, but that the extent to which information must be maintained or manipulated mediates how various substrates are involved. Both of

these factors will confound the interpretation of data in experiments interrogating mechanisms of STM, even those that succeed in probing the information content of non-spiking neural states. They also imply that neural states will need to be compared across tasks of varied cognitive demands. If STM involves potentially many substrates, with the specific balance mediated by needs for maintenance and manipulation of information, rigorous experiments to elucidate STM mechanisms will need to train animals on several tasks and measure the relevant substrates for each. All of this is to say that, even with current and foreseeable advances in approaches for monitoring neural population activity during behavior, unraveling the key mechanisms and computations of STM will be difficult using animal models alone. How might investigators refine their hypotheses about WM to design experiments in light of these factors (e.g., without incurring prohibitive cost)? Here, we highlight recent evidence that points to a powerful new option: machine-learning-based RNNs.

Machine Learning Models of Activity-Silent STM

The studies discussed so far have provided strong evidence that neural ‘hidden states’, like STSP, can maintain information for STM. However, further progress in our understanding is hindered by the fact that hidden states are, as the name suggests, effectively hidden: it is currently extremely challenging to directly measure synaptic efficacies in awake, behaving animals such as mice and nonhuman primates. One promising alternative is to model STM and WM using machine-learning-based RNNs. RNN models can provide insights into the putative circuit functions supporting cognitive processes that would otherwise be unattainable through direct experimental measurement. By training RNNs and experimental subjects on the same tasks, researchers can perform analyses in parallel, a process facilitated by the fact that the neural activity and connectivity of the RNN are fully known. Crucially, RNNs trained on cognitive tasks can exhibit single-cell and population dynamics that mirror actual neural activity patterns observed in neurophysiological recordings [65–69]. The same is also true in machine-learning-based feedforward neural networks trained to classify images, whose patterns of neural responses are often very similar to those recorded from corresponding levels of the ventral visual hierarchy [70,71].

A recent study from our group sought to understand why the strength of persistent delay activity varied markedly between tasks and, for tasks with low levels of persistent activity, where and how memoranda are maintained [23]. To do so, biologically inspired RNNs comprising excitatory- and inhibitory-like neurons and synapses governed by STSP (Figure 5A) were trained to solve a variety of STM- and WM-based tasks involving comparisons between sequentially presented sample and test stimuli, separated by a delay period (Figure 5B). We found that networks were able to solve tasks requiring only passive STM, such as the delayed match-to-sample task (Figure 5C, top panel), in an ‘activity-silent’ manner (without delay-period persistent activity) by maintaining stimulus information in short-term changes in synaptic efficacies due to STSP. By contrast, tasks that did require manipulation, such as a delayed match-to-rotated-sample task (Figure 5C, bottom panel), always required some level of persistent activity. Interestingly, when we compared the level of persistent activity across a variety of different tasks, we found that tasks requiring a greater amount of information manipulation also required greater persistent activity (Figure 5D).

Furthermore, analysis of these networks revealed that the stimulus information encoded in STSP does not necessarily need to be reloaded back into neural activity to generate a behavioral response. Rather, STSP can prospectively encode the sample stimulus, altering network dynamics so that it correctly responds to the upcoming test stimulus. This is a nice example of a synergistic interaction between the activity-silent substrate (STSP) and the more active neural encoding

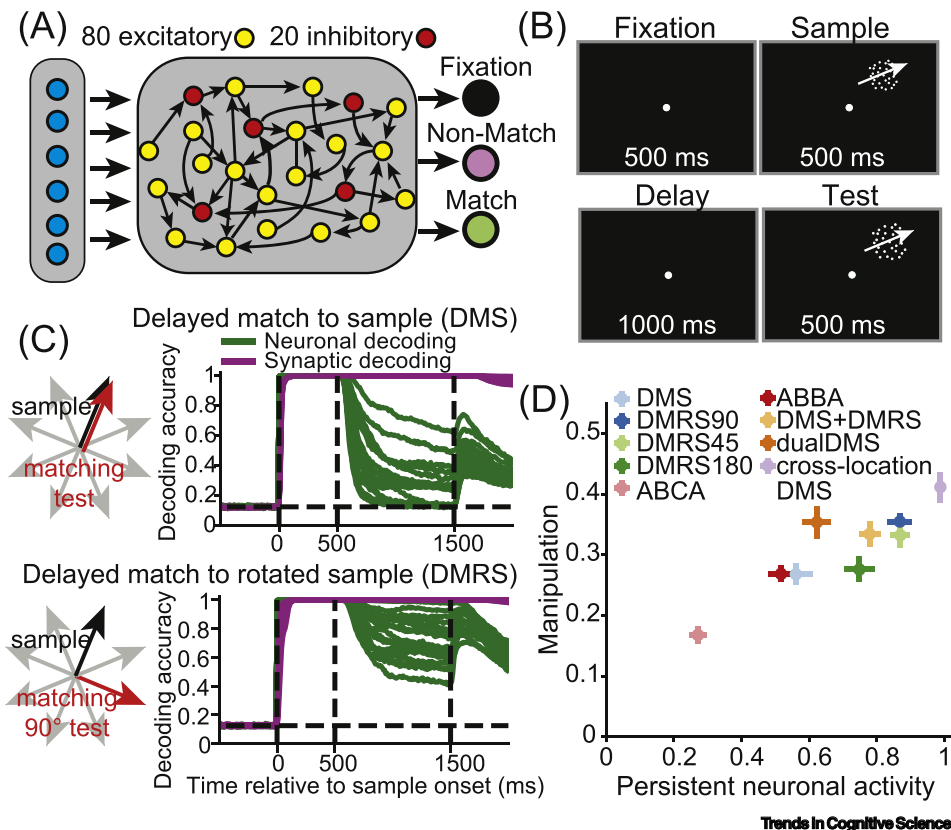


Figure 5. Short-Term Memory (STM) and Working Memory (WM) in Recurrent Neural Networks (RNNs). Studying STM and WM with biologically inspired RNNs. (A) The rate-based model comprised 24 motion-direction-tuned neurons (blue) projecting onto 80 excitatory (yellow) and 20 inhibitory (red) recurrently connected neurons. The 80 excitatory neurons projected onto three decision neurons. Recurrent connections were governed by short-term synaptic plasticity (STSP). (B) A 500-ms fixation period was followed by a 500-ms sample motion-direction stimulus, followed by a 1000-ms delay period and finally a 500-ms test stimulus. (C) Top panel: In the delayed match-to-sample (DMS) task, the network had to indicate whether the test motion direction (red arrow) matched the sample motion direction (black arrow). Sample decoding accuracy, calculated using neuronal activity (green curves) and synaptic efficacy (magenta curves) for $n = 20$ networks. Each curve represents the decoding accuracy for one network. Bottom panel: In the delayed match-to-rotated-sample (DMRS) task, the network had to indicate whether the test motion direction was 90° clockwise from the sample motion direction. (D) Scatterplot shows the level of persistent neuronal activity, measured as the neuronal decoding accuracy during the last 100ms of the delay (x-axis), versus the level of manipulation (y-axis) across nine different tasks (indicated by colored crosses). Adapted from [23].

of the test stimulus. The interaction between various neural substrates and mechanisms undoubtedly occurs in the brain, although this has been relatively unexplored. Future modeling and experimental work will hopefully uncover whether synergistic interactions between substrates is an important feature in STM and WM.

There are several take-away points from this study that are relevant for the work described earlier in this review. First, while the maintenance of information can happen without persistent activity, the manipulation of information does require persistent activity, consistent with the experimental findings described above [38]. In one of the studies from our group mentioned above, stimulus-selective persistent activity was weak in the posterior parietal cortex area LIP during a delayed motion-direction-matching task [14,15] but became robust when the subjects were trained to transform the motion direction into an abstract category representation [14]. We also found

that persistent activity was weak to absent for RNN models trained on the delayed direction-matching task but became robust when trained on the delayed category-matching task [23]. Other tasks that do not explicitly require the direct manipulation of memory contents might require other forms of manipulation (e.g., manipulation of the remembered sensory stimulus to generate motor commands). This might explain why memory-guided saccade and reaching tasks seem to require persistent activity, as in studies of prospective/retrospective coding that dissociate between the location of a stimulus and the target saccade [34]. Finally, information can be prospectively encoded in STSP, consistent with recent human behavioral evidence that STM and WM encoding can be prospective in nature [72]. The common thread across all points is that STM is highly task dependent and that no single task can be used to unravel its mechanisms.

This task dependence is also the conclusion of another RNN-based modeling study, which showed that even simple changes to the task, such as switching between a fixed delay period and one with a random duration, can affect the nature of persistent activity [73]. The authors of this study showed further that various changes to the network structure, such as varying the connectivity between neurons, can produce different forms of persistent activity, suggesting that different network solutions to the same problem can also result in variations in STM-related activity. Thus, the variations in persistent activity observed in past studies might reflect differences in network structure and connectivity between subjects, in addition to differences in task demands or training history.

Although these machine-learning-based modeling studies cannot replace experimental work, they can prove indispensable for identifying the putative circuit mechanisms and dynamics underlying STM, WM, and other cognitive functions, particularly when obtaining the necessary experimental data is difficult or impossible. Thus, these studies can serve as a complement to experimental work, allowing researchers to rapidly generate hypotheses about neural mechanisms that can be tested when technology better allows experimental verification.

Concluding Remarks

Our understanding of the neural mechanisms supporting STM and WM, historically and successfully centered around persistent spiking, has begun to expand in new directions. This expansion has illuminated the mapping between cognitive processes and their underlying neural mechanisms. To a first approximation, passive, non-spiking mechanisms (mostly) suffice when the contents of memory need only to be maintained over short durations and active spiking mechanisms are required when these contents must be processed to generate a behavior or to bridge delay periods much longer than the time constant of STSP. This is not to suggest that passive maintenance exclusively relies on non-spiking mechanisms or that active manipulation exclusively relies on spiking mechanisms, but rather that the distribution of neural substrates engaged will shift depending on the cognitive demands of the task. While previous neurophysiological studies of STM and WM have observed this heterogeneity in the neural response [10–16], its dependence on task demands has only recently been emphasized [38,39].

At least two problems complicate our path to a holistic understanding of STM. First, both the spiking and (especially) the non-spiking aspects of large-scale neural circuit activity are difficult to measure; without these data, any picture of STM will remain incomplete. Second, the balance of substrates engaged for STM seems to depend on task demands, a confound avoidable only by testing multiple tasks. Training animals to perform even one STM task and collecting neurophysiological data requires significant expenditure of time and resources, so the process of disentangling task influence through animal experiments alone seems likely to scale poorly.

Outstanding Questions

How can we identify which non-spiking neural substrates are used by the brain to maintain and manipulate information, of the many possibilities active over different timescales? Is information primarily concentrated in a small number of core substrates or is it distributed across a wide range of different biophysical processes? What factors influence how information for STM engages different substrates?

How do the various neural substrates involved in STM interact and how do we model these interactions?

Can spiking neural networks (SNNs) that faithfully reproduce the spiking dynamics of neurons provide greater insights into the neural mechanisms underlying STM *in vivo*? How can we train machine-learning-based SNNs?

Does there exist an underlying principle that governs how memoranda are maintained across a distributed set of neural substrates, operating across a diverse set of timescales?

What are the potential benefits and drawbacks of maintaining information across a distributed set of neural substrates? Does it primarily confer robust maintenance of information across a wider range of timescales or does it endow us with additional computational advantages?

Does the distributed set of neural substrates involved in STM play a role analogous to the set of coupled neural processes that support long-term memory? How do the computational needs of memory over short timescales differ from those required over long timescales?

What role do non-spiking neuronal substrates play outside the cognitive context of STM?

To bridge this gap, we and others have begun to explore RNNs as viable model organisms, ones whose attributes could facilitate a role complementary to that filled by traditional animal systems. RNNs require few resources to train and all knowledge of the neural activity and synaptic connectivity is fully known, benefits that no vertebrate animal model enjoys. Some of this work is already occurring as part of a feedback loop between neuroscience and artificial intelligence [74,75]. In one direction, insights about the brain are increasingly used to generate advances in artificial neural networks [67,76–80]. In the other direction, artificial neural networks can be used to generate novel hypotheses about the mechanisms underlying various cognitive functions [23,67,70,80–82].

However, the extent to which principles active in RNNs bear on biological brains remains to be elucidated. Future work will need to explore which comparisons between networks *in vivo* and *in silico* ring especially true, as well as where the analogy falls apart. Given that neurons *in vivo* represent information through spiking as opposed to an analog representation, there is a clear need to develop machine-learning algorithms to train spiking neural networks to enable a more faithful comparison with biological networks of neurons. This is especially relevant given that many processes *in vivo*, ranging from oscillatory activity to channel dynamics, are inherently tied to the spiking nature of biological neurons.

The experimental and modeling studies we have described here begin to paint the contours of a picture of STM and WM – enough to suggest where richer detail should be, but not to supply it. Filling out those features that remain – in particular, how different neural substrates subserve various STM processes and how these substrates interact – will require an integrated approach that combines the techniques described in this review. Together, well-controlled behavioral assays, new approaches to record from neural various substrates, and biologically inspired network models will furnish a more complete understanding of the diverse processes underlying STM (see Outstanding Questions).

References

1. Baddeley, A.D. and Hitch, G. (1974) Working memory. In *Psychology of Learning and Motivation* (Vol. 8) (Bower, G.H., ed.), pp. 47–89, Academic Press
2. Baddeley, A. (1992) Working memory. *Science* 255, 556–559
3. Fukuda, K. *et al.* (2010) Quantity, not quality: the relationship between fluid intelligence and working memory capacity. *Psychon. Bull. Rev.* 17, 673–679
4. Funahashi, S. *et al.* (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* 61, 331–349
5. Chafee, M.V. and Goldman-Rakic, P.S. (1998) Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *J. Neurophysiol.* 79, 2919–2940
6. Rainer, G. *et al.* (1998) Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature* 393, 577–579
7. Romo, R. *et al.* (1999) Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* 399, 470–473
8. Stokes, M.G. (2015) 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn. Sci.* 19, 394–405
9. Lundqvist, M. *et al.* (2018) Working memory: delay activity, yes! Persistent activity? Maybe not. *J. Neurosci.* 38, 7013–7019
10. 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
11. Sreenivasan, K.K. *et al.* (2014) Revisiting the role of persistent neural activity during working memory. *Trends Cogn. Sci.* 18, 82–89
12. Lee, S.-H. *et al.* (2013) Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nat. Neurosci.* 16, 997–999
13. Lara, A.H. and Wallis, J.D. (2014) Executive control processes underlying multi-item working memory. *Nat. Neurosci.* 17, 876–883
14. Sarma, A. *et al.* (2016) Task-specific versus generalized mnemonic representations in parietal and prefrontal cortices. *Nat. Neurosci.* 19, 143–149
15. Masse, N.Y. *et al.* (2017) Mnemonic encoding and cortical organization in parietal and prefrontal cortices. *J. Neurosci.* 37, 6098–6112
16. 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
17. Emrich, S.M. *et al.* (2013) Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. *J. Neurosci.* 33, 6516–6523
18. 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
19. Mongillo, G. *et al.* (2008) Synaptic theory of working memory. *Science* 319, 1543–1546
20. Rose, N.S. *et al.* (2016) Reactivation of latent working memories with transcranial magnetic stimulation. *Science* 354, 1136–1139
21. Wolff, M.J. *et al.* (2017) Dynamic hidden states underlying working-memory-guided behavior. *Nat. Neurosci.* 20, 864–871
22. Sprague, T.C. *et al.* (2016) Restoring latent visual working memory representations in human cortex. *Neuron* 91, 694–707
23. Masse, N.Y. *et al.* (2019) Circuit mechanisms for the maintenance and manipulation of information in working memory. *Nat. Neurosci.* 22, 1159
24. Cowan, N. (2008) What are the differences between long-term, short-term, and working memory? In *Progress in Brain Research*, pp. 323–338, Elsevier
25. Aben, B. *et al.* (2012) About the distinction between working memory and short-term memory. *Front. Psychol.* 3, 301

26. Cowan, N. (2017) The many faces of working memory and short-term storage. *Psychon. Bull. Rev.* 24, 1158–1170
27. Hopfield, J.J. (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U. S. A.* 79, 2554–2558
28. Inagaki, H.K. *et al.* (2019) Discrete attractor dynamics underlies persistent activity in the frontal cortex. *Nature* 566, 212
29. Li, N. *et al.* (2016) Robust neuronal dynamics in premotor cortex during motor planning. *Nature* 532, 459–464
30. Inagaki, H.K. *et al.* (2018) Low-dimensional and monotonic preparatory activity in mouse anterior lateral motor cortex. *J. Neurosci.* 38, 4163–4185
31. 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
32. Brody, C.D. *et al.* (2003) Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cereb. Cortex* 13, 1196–1207
33. Shafi, M. *et al.* (2007) Variability in neuronal activity in primate cortex during working memory tasks. *Neuroscience* 146, 1082–1108
34. Takeda, K. and Funahashi, S. (2004) Population vector analysis of primate prefrontal activity during spatial working memory. *Cereb. Cortex* 14, 1328–1339
35. Murray, J.D. *et al.* (2017) Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proc. Natl. Acad. Sci. U. S. A.* 114, 394–399
36. Wasmuht, D.F. *et al.* (2018) Intrinsic neuronal dynamics predict distinct functional roles during working memory. *Nat. Commun.* 9, 3499
37. Mendoza-Halliday, D. *et al.* (2014) Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nat. Neurosci.* 17, 1255–1262
38. Trübetschek, D. *et al.* (2019) Probing the limits of activity-silent non-conscious working memory. *Proc. Natl. Acad. Sci. U. S. A.* 116, 14358–14367
39. Christophel, T.B. *et al.* (2018) Cortical specialization for attended versus unattended working memory. *Nat. Neurosci.* 21, 494–496
40. Kulkushkin, N.V. and Carew, T.J. (2017) Memory takes time. *Neuron* 95, 259–279
41. Szatmáry, B. and Izhikevich, E.M. (2010) Spike-timing theory of working memory. *PLoS Comput. Biol.* 6, e1000879
42. Fiebig, F. and Lansner, A. (2017) A spiking working memory model based on Hebbian short-term potentiation. *J. Neurosci.* 37, 83–96
43. Lisman, J.E. *et al.* (1998) A role for NMDA-receptor channels in working memory. *Nat. Neurosci.* 1, 273–275
44. Wang, X.-J. (1999) Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. *J. Neurosci.* 19, 9587–9603
45. Wang, X.-J. (2001) Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci.* 24, 455–463
46. Wang, M. *et al.* (2013) NMDA receptors subserve persistent neuronal firing during working memory in dorsolateral prefrontal cortex. *Neuron* 77, 736–749
47. Hasselmo, M.E. (2006) The role of acetylcholine in learning and memory. *Curr. Opin. Neurobiol.* 16, 710–715
48. Qi, X.-L. *et al.* (2019) Nucleus basalis stimulation stabilizes attractor networks and enhances task representation in prefrontal cortex. *bioRxiv*. Published online June 18, 2019. <https://doi.org/10.1101/674465>
49. Murphy, B.L. *et al.* (1996) Increased dopamine turnover in the prefrontal cortex impairs spatial working memory performance in rats and monkeys. *Proc. Natl. Acad. Sci. U. S. A.* 93, 1325–1329
50. Vijayraghavan, S. *et al.* (2007) Inverted-U dopamine D1 receptor actions on prefrontal neurons engaged in working memory. *Nat. Neurosci.* 10, 376–384
51. Wang, M. *et al.* (2007) $\alpha 2A$ -Adrenoceptors strengthen working memory networks by inhibiting cAMP-HCN channel signaling in prefrontal cortex. *Cell* 129, 397–410
52. Sawaguchi, T. and Goldman-Rakic, P.S. (1991) D1 dopamine receptors in prefrontal cortex: involvement in working memory. *Science* 251, 947–950
53. Ott, T. and Nieder, A. (2019) Dopamine and cognitive control in prefrontal cortex. *Trends Cogn. Sci.* 23, 213–234
54. Serences, J.T. (2016) Neural mechanisms of information storage in visual short-term memory. *Vis. Res.* 128, 53–67
55. Christophel, T.B. *et al.* (2017) The distributed nature of working memory. *Trends Cogn. Sci.* 21, 111–124
56. Harrison, S.A. and Tong, F. (2009) Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458, 632–635
57. Serences, J.T. *et al.* (2009) Stimulus-specific delay activity in human primary visual cortex. *Psychol. Sci.* 20, 207–214
58. Ranganath, C. (2004) Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. *J. Neurosci.* 24, 3917–3925
59. Kornblith, S. *et al.* (2017) Persistent single-neuron activity during working memory in the human medial temporal lobe. *Curr. Biol.* 27, 1026–1032
60. Kamiński, J. *et al.* (2017) Persistently active neurons in human medial frontal and medial temporal lobe support working memory. *Nat. Neurosci.* 20, 590–601
61. Hallock, H.L. *et al.* (2016) Ventral midline thalamus is critical for hippocampal–prefrontal synchrony and spatial working memory. *J. Neurosci.* 36, 8372–8389
62. Funahashi, S. (2013) Thalamic mediodorsal nucleus and its participation in spatial working memory processes: comparison with the prefrontal cortex. *Front. Syst. Neurosci.* 7, 36
63. Roux, F. and Uhlhaas, P.J. (2014) Working memory and neural oscillations: alpha-gamma versus theta-gamma codes for distinct WM information? *Trends Cogn. Sci.* 18, 16–25
64. Miller, E.K. *et al.* (2018) Working memory 2.0. *Neuron* 100, 463–475
65. Cueva, C.J. *et al.* (2019) Low dimensional dynamics for working memory and time encoding. *bioRxiv*. Published online January 31, 2019. <https://doi.org/10.1101/504936>
66. Chaisangmongkon, W. *et al.* (2017) Computing by robust transience: how the fronto-parietal network performs sequential, category-based decisions. *Neuron* 93, 1504–1517.e4
67. Banino, A. *et al.* (2018) Vector-based navigation using grid-like representations in artificial agents. *Nature* 557, 429–433
68. Wang, J. *et al.* (2018) Flexible timing by temporal scaling of cortical responses. *Nat. Neurosci.* 21, 102–110
69. Cueva, C.J. and Wei, X.-X. (2018) Emergence of grid-like representations by training recurrent neural networks to perform spatial localization. In *Proceedings of the International Conference on Learning Representations (ICLR)*, ICLR
70. Yamins, D.L.K. *et al.* (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* 111, 8619–8624
71. Kar, K. *et al.* (2019) Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nat. Neurosci.* 22, 974–983
72. Printzlau, F. *et al.* (2019) Prospective task knowledge improves working memory-guided behaviour. *PsyArXiv* Published online July 18, 2019. <http://dx.doi.org/10.31234/osf.io/qhku6>
73. Orhan, A.E. and Ma, W.J. (2019) A diverse range of factors affect the nature of neural representations underlying short-term memory. *Nat. Neurosci.* 22, 275
74. Hassabis, D. *et al.* (2017) Neuroscience-inspired artificial intelligence. *Neuron* 95, 245–258
75. Richards, B.A. *et al.* (2019) A deep learning framework for neuroscience. *Nat. Neurosci.* 22, 1761–1770
76. Krizhevsky, A. *et al.* (2017) ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60, 84–90
77. Graves, A. *et al.* (2016) Hybrid computing using a neural network with dynamic external memory. *Nature* 538, 471–476
78. Kirkpatrick, J. *et al.* (2017) Overcoming catastrophic forgetting in neural networks. *Proc. Natl. Acad. Sci. U. S. A.* 114, 3521–3526
79. Masse, N.Y. *et al.* (2018) Alleviating catastrophic forgetting using context-dependent gating and synaptic stabilization. *Proc. Natl. Acad. Sci. U. S. A.* 115, E10467–E10475
80. Wang, J.X. *et al.* (2018) Prefrontal cortex as a meta-reinforcement learning system. *Nat. Neurosci.* 21, 860–868
81. Nicola, W. and Clopath, C. (2019) A diversity of interneurons and Hebbian plasticity facilitate rapid compressible learning in the hippocampus. *Nat. Neurosci.* 22, 1168–1181
82. Yang, G.R. *et al.* (2019) Task representations in neural networks trained to perform many cognitive tasks. *Nat. Neurosci.* 22, 297–306

83. Seung, H.S. (1996) How the brain keeps the eyes still. *Proc. Natl. Acad. Sci. U. S. A.* 93, 13339–13344
84. Mazor, O. and Laurent, G. (2005) Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron* 48, 661–673
85. Churchland, M.M. *et al.* (2012) Neural population dynamics during reaching. *Nature* 487, 51–56
86. Harvey, C.D. *et al.* (2012) Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* 484, 62–68
87. Mante, V. *et al.* (2013) Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503, 78–84
88. Sadtler, P.T. *et al.* (2014) Neural constraints on learning. *Nature* 512, 423–426
89. Miller, E. *et al.* (1993) Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *J. Neurosci.* 13, 1460–1478
90. Bisley, J.W. *et al.* (2004) Activity of neurons in cortical area MT during a memory for motion task. *J. Neurophysiol.* 91, 286–300
91. Offen, S. *et al.* (2009) The role of early visual cortex in visual short-term memory and visual attention. *Vis. Res.* 49, 1352–1362
92. Rademaker, R.L. *et al.* (2019) Coexisting representations of sensory and mnemonic information in human visual cortex. *Nat. Neurosci.* 22, 1336–1344
93. Foster, J.J. *et al.* (2016) The topography of alpha-band activity tracks the content of spatial working memory. *J. Neurophysiol.* 115, 168–177
94. Logothetis, N.K. *et al.* (2001) Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412, 150–157
95. Lee, H. *et al.* (2005) Phase locking of single neuron activity to theta oscillations during working memory in monkey extrastriate visual cortex. *Neuron* 45, 147–156
96. Buschman, T.J. and Miller, E.K. (2007) Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* 315, 1860–1862
97. Gregoriou, G.G. *et al.* (2009) High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science* 324, 1207–1210
98. Bahmani, Z. *et al.* (2018) Working memory enhances cortical representations via spatially specific coordination of spike times. *Neuron* 97, 967–979.e6
99. Ester, E.F. *et al.* (2015) Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. *Neuron* 87, 893–905