Ann. N.Y. Acad. Sci. ISSN 0077-89

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Special Issue: *The Year in Cognitive Neuroscience*

REVIEW

Between persistently active and activity-silent frameworks: novel vistas on the cellular basis of working memory

Jan Kamiński[1,2] and Ueli Rutishauser[1,3,4,2]

[1]Department of Neurosurgery, Cedars-Sinai Medical Center, Los Angeles, California. [2]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California. [3]Department of Neurology, Cedars-Sinai Medical Center, Los Angeles, California. [4]Center for Neural Science and Medicine, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, California.

Address for correspondence: Jan Kamiński, Department of Neurosurgery, Cedars-Sinai Medical Center, 127 S. S San Vicente Blvd, Los Angeles, CA 90048—1804. jan.kaminski@cshs.org

Recent work has revealed important new discoveries on the cellular mechanisms of working memory (WM). These findings have motivated several seemingly conflicting theories on the mechanisms of short-term memory maintenance. Here, we summarize the key insights gained from these new experiments and critically evaluate them in light of three hypotheses: classical persistent activity, activity-silent, and dynamic coding. The experiments discussed include the first direct demonstration of persistently active neurons in the human medial temporal lobe that form static attractors with relevance to WM, single-neuron recordings in the macaque prefrontal cortex that show evidence for both persistent and more dynamic types of WM representations, and noninvasive neuroimaging in humans that argues for activity-silent representations. A key insight that emerges from these new results is that there are several neural mechanisms that support the maintenance of information in WM. Finally, based on established cognitive theories of WM, we propose a coherent model that encompasses these seemingly contradictory results. We propose that the three neuronal mechanisms of persistent activity, activity-silent, and dynamic coding map well onto the cognitive levels of information processing (within focus of attention, activated long-term memory, and central executive) that Cowan's WM model proposes.

Keywords: working memory; single-neuron recordings; persistent activity; dynamic coding; static coding; attractors

Introduction

Working memory (WM, see Table 1 for a list of acronyms) is the capacity to hold and manipulate information in mind.[1] WM is a fundamental cognitive function that allows us to execute complex tasks in a constantly changing environment.[1] Recent years have brought substantial advancement in the field of WM and this has driven an emergence of new hypotheses regarding the neuronal mechanism of how we hold and manipulate items in mind. The goal of this review is to describe these new discoveries and consider them in the context of established cognitive frameworks of WM. Our focus here is on new discoveries made at the single neuron level, which provides an unprecedented opportunity to directly observe the mechanisms that support WM.

Here, we summarize findings that together show that items in WM can be stored using two kinds of mechanisms: one decodable from neural activity measured by electrophysiological or metabolic means and one that is not (Fig. 1). Next, we examine the newly emerging view that there are two forms of active representations that maintain working memories: stable, persistently active neurons and dynamically active neurons. We conclude by proposing how these three different types of cellular mechanisms for maintaining information fit into the cognitive frameworks of WM.

Single-cell evidence for persistent activity

Almost 50 years ago, scientists for the first time observed neurons that continue to fire during

doi: 10.1111/nyas.14213

Table 1. Acronyms

ALM	Anterior lateral motor cortex
BOLD-fMRI	Blood-oxygen-level-dependent functional magnetic resonance imaging
EEG	Electroencephalogram
FEF	Frontal eye field
IPS	Intraparietal sulcus
LFP	Local field potentials
LTM	Long-term memory
MTL	Medial temporal lobe
PA	Persistent activity (of neurons)
PCA	Principal component analysis
PFC	Prefrontal cortex
STSP	Short-term synaptic plasticity
WM	Working memory

the maintenance period after the end of stimulus presentation.[2] Such activity can last for many seconds and is stimulus specific: cells continue to fire only if their preferred stimulus (which typically is a specific sensory input or the direction of an instructed motor movement to be executed later) is held in WM. This pattern of activity has become known as *persistent activity* (PA, also called delay activity) because it outlasts the time of stimulus presentation. Subsequently, PA has become the central element in theories of the neuronal mechanism of WM.[3] To date, signatures of PA have been observed at the single-cell level in many brain areas, species, and experimental paradigms.[4–16]

In addition to much work in animal models, recent work has revealed direct evidence for PA and its relevance to WM in humans.[17,18] This work was performed as part of invasive monitoring for seizure localization, a clinical procedure for which depth electrodes with embedded microwires are implanted in human patients suffering from drug-resistant epilepsy.[19] Using this method, the electrical activity of individual neurons is recorded while patients perform cognitive tasks.[20] The first direct evidence for WM relevant PA in human neurons[17,18] shows that highly selective "concept" cells in the human medial temporal lobe (MTL) can remain active for several seconds if the concept that activates these neurons is held in WM (Fig. 2A). Concept cells are a type of cell whose properties have been studied extensively in humans and are therefore relatively well understood.[21,22] However, so far, concept cells have been viewed as representing aspects of declarative memories and not those of WM.[23] Persistently active concept cells were found in multiple areas of the MTL, including the hippocampus, amygdala, parahippocampal cortex, and entorhinal cortex. The strength of this activity predicted behavior and scaled with WM load.[17] In addition, nonstimulus-specific PA in the MTL has also been observed.[17,24] Together, this body of work reveals evidence for persistently active cells in the human MTL whose activity is related to WM.

Attractors as a framework to study PA

What are the mechanisms that give rise to sustained neuronal activity? One possibility is cell autonomous mechanisms, which exist in certain specialized cells,[25,26] including in humans.[27] Another possibility is at the network level, facilitated by recurrent synaptic connections. By and large, current theories assume that the PA that gives rise to WM is due to network-level effects.[26,28] However, in most experiments conducted so far (see Ref. 7 for a notable exception), it remains unknown whether the observed PA is indeed dependent on network-level interactions rather than cell-autonomous mechanisms. The contribution of cell-autonomous PA to WM thus remains an important open question.

A useful theoretical framework to conceptualize network-level PA is that of attractors, which are stable patterns of neural activity that are maintained through recurrent excitation.[29,30] Each possible pattern constitutes a different possible item held in WM. There are two different classes of attractor models that are typically considered in this context: continuous attractor networks, which have a continuum of stable states ideal for encoding analog variables, and discrete attractors networks that have a countable number of possible discrete states that compete with each other. Recent single-neuron studies in several different species (mice, rhesus, and humans) have provided direct experimental evidence for the presence of such attractors during WM maintenance and their relation to behavior. We will next summarize these key findings.

Recently, we used demixed PCA to assess the population dynamics associated with the identity of stimuli held in WM.[17,31] This revealed that during WM maintenance, the speed of the neuronal trajectories in the dimensions associated with stimulus identity was low (comparable to the speed present at

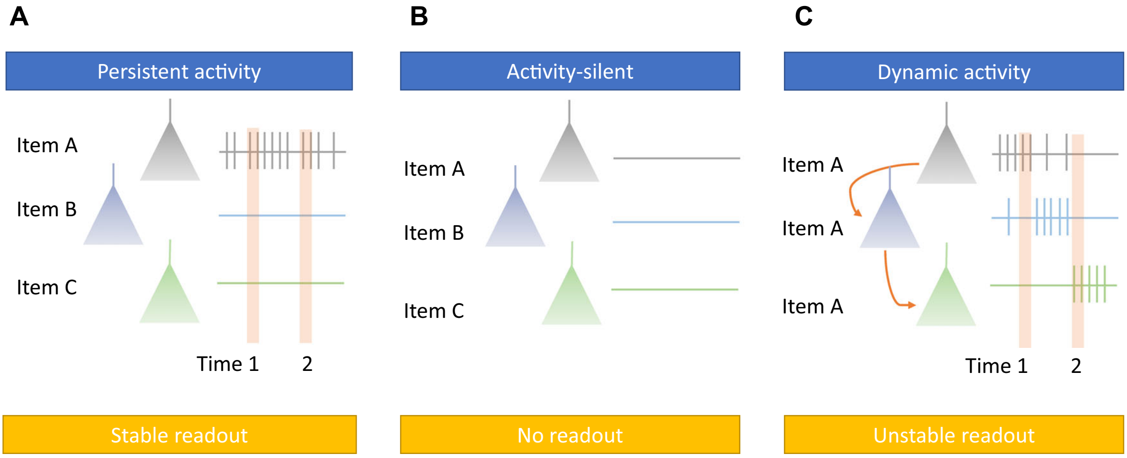

Figure 1. Summary of theories of the neuronal mechanisms supporting working memory. (A) The persistent activity framework proposes that memoranda are maintained by the sustained firing of stimulus-specific groups of neurons. A decoder trained in one period of time should be able to decode information at a different point of time ("1" versus "2" periods indicated). (B) The activity-silent framework proposes that information held in WM is not visible by observing the activity of individual neurons. (C) The dynamic coding framework proposes that the neurons carrying information about a specific item change as function of time relative to the onset of the maintenance period. For example, some neurons encode the identity of an item only at a specific period of time. In the figure, three neurons are shown, all of which represent item A, but during different periods of time, with some neurons "ramping down" their activity (top), whereas others firing only during specific periods of time. A decoder trained at one point of time will thus not generalize to a different point of time ("1" versus "2" periods indicated).

baseline). At the same time, we found that the distance between the trajectories associated with different items was high. This suggests that neuronal activity was pulled to a particular area in state space and then remained there, which is the definition of a discrete attractor (Fig. 2B).[29] These attractors were behaviorally relevant: the distance of the neuronal trajectory in a given trial to the center of the attractor was correlated with later accuracy and reaction time for a test stimulus. This revealed that in trials when activity drifted away from the center of the attractor, the quality of the memory decreased.

Another study has shown evidence for continuous "bump attractors" by revealing a direct relation between attractor dynamics and behavior.[32] A bump attractor is a type of continuous attractor, where activity forms a bell-shape like pattern of activity that is centered on the value of a currently maintained memorandum. Monkeys performed an oculomotor spatial WM task that required memorizing one of eight possible spatial locations. During maintenance, the authors found stimulus-specific PA in the prefrontal cortex. The fluctuations in activity of these neurons correlated trial-by-trial with inaccuracies of the behavioral response (Fig. 2C). These fluctuations were such that neurons that encoded positions adjacent to the currently cued location increased their firing rate proportionally to the behavioral bias toward the position preferred by a given cell. This linear relationship could be explained by continuous, but not by discrete attractor dynamics,[32] thereby revealing experimental evidence for continuous attractors relevant for WM at the single-cell level.

Attractor dynamics were also closely examined in a study in which mice needed to maintain information about the position of a reward (left or right).[7] First, extracellular recordings in the anterior lateral motor cortex (ALM) revealed evidence for strong memory content–specific PA that formed attractors in state space. Critically, the authors showed that the PA in this case was not due to cell-autonomous PA. To achieve this, they hyperpolarized persistently active cells using whole-cell recordings. This abolished spiking activity but not other signs of persistent synaptic activation as measured by the subthreshold membrane potential, thereby revealing a network-level origin of PA in ALM.[33] This is, to our knowledge, the strongest *in-vivo* demonstration of the network origin of PA during WM thus far. Second, strong optogenetic inhibition of ALM cells abolished the ability of mice to maintain

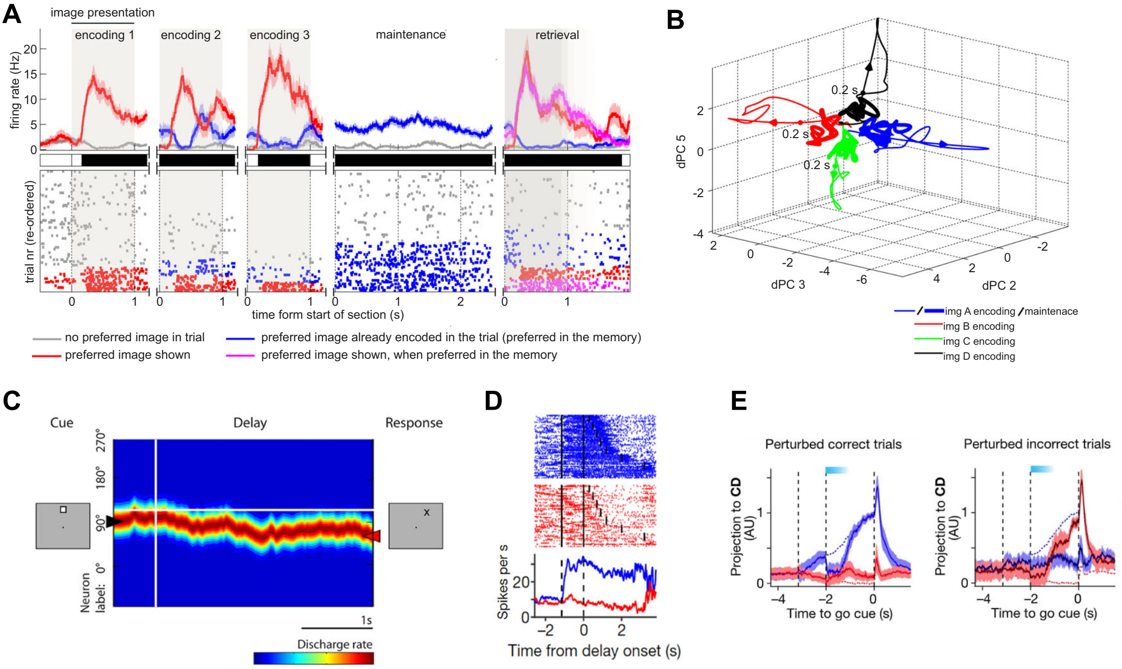

Figure 2. Persistent activity represented by attractor dynamic. (A) Example of a persistently active concept cell recorded in the human amygdala. Subjects memorized up to three images presented sequentially (encoding 1–3). Top: post stimulus time histogram. Middle: periods of significance (black) between the preferred versus nonpreferred stimuli of this cell. Bottom: raster plot of trials reordered according to condition. During maintenance, the activity of this cell is characterized by sustained activity only when the preferred image of the cell is held in memory (blue) but not when other stimuli are held in memory (gray). Adapted from Ref. 17. (B) Trajectories in neural state space formed by a population of persistently active concept cells in the human MTL. Trajectories are projected into the 3D space formed by the three demixed principal components (dPCs) associated with picture identity. Periods of time shown are encoding (thin line) and maintenance (thick line). Colors mark different images. Note how during maintenance, activity settles at points in space that separate by memory content (attractors). Adapted from Ref. 17. (C) Neuron whose activity is indicative of a continuous attractor during a delayed oculomotor task in rhesus monkey. Firing rate of stimulus-selective neurons is sorted vertically according to preferred location. Note how activity drifts away from the initial position (left) as time progresses. This drift predicts behavioral errors (right). Adapted from Ref. 50. (D) Persistent activity recorded in mouse ALM during a task with variable delay durations. Blue color marks prefered location. (E) Population activity in mouse ALM shows characteristic of a discrete attractor. Shown is a projection of the population activity onto the axis that maximally distinguishing between the two possible conditions (left or right). Blue color (left panel) denotes correct lick-right trials, and red denotes correct lick-left trials. Dark blue and dark red (right panel) denote incorrect lick-right and lick-left trials, respectively. Dashed lines denote trajectories of unperturbed correct trials, whereas solid lines denote perturbed trials. Light blue band on the top shows time of photoinhibition. Note how the neural activity after offset of inhibition is pulled toward one of the two possible trajectories, as expected from an attractor. Adapted from Ref. 7.

information in WM, showing the necessity of these cells for information maintenance. Weak inhibition, on the other hand, revealed a remarkable phenomenon: in some trials, such inhibition led to an error, whereas in others it did not. In error trials, the PA following offset of the inhibition resembled that of the opposite direction (that was not cued), whereas in correct trials, activity returned to the cued direction. Thus, the pattern of neuronal activity was attracted toward one of two discrete states (attractors, Fig. 2E) to which activity returned after transient disruption. Together, this study shows compelling evidence for network-level PA that forms discrete attractors.

The study of Inagaki *et al.*[7] revealed a critical difference between when the maintenance duration was of fixed versus randomized duration: stable PA was observed only in the latter (Fig. 2D). In contrast, for fixed durations, stimulus selective delay-period activity was characterized by slow, ramping activity. While not classically expected from discrete attractor dynamics, slow ramping can be incorporated into discrete attractor networks by making network activity move slowly to the attractor center.[8] Indeed, in their analysis, Inagaki *et al.* showed that all other characteristics of neuronal activity, such as trajectory recovery and the bimodal distribution of endpoints, showed that these dynamics were nevertheless compatible with discrete attractor states. Similar ramping activity for fixed but not variable length delay periods has also been observed in macaque PFC.[34] These data point out an important insight: in the presence of predictable delay-period durations, PA can also reflect aspects of time progression. It is therefore critical to randomize maintenance durations to differentiate these different modes of delay-period activity. Note, in our human experiments,[17] delay-period duration was randomized by ±150 ms, which has much less variance than the seconds in the rhesus and mice tasks discussed above.[7,34]

In the work discussed above, two different types of attractors were identified (continuous versus discrete). However, this difference could be a function of the different task designs. Whereas in the study of Inagaki *et al.*, animals made a choice from two well-defined, discreet positions, in the study of Wimmer *et al.*, monkeys needed to remember a cue position in continuous space. These two tasks thus had different demands, which might have resulted in different network dynamics. Overall, the summarized work shows that attractor network dynamics are a powerful framework to test the characteristics of delay-period activity.

Interactions between WM and long-term memory

Cowan proposed the embedded-processes organization of WM, which can be divided into three levels: (1) long-term memory (LTM), (2) part of LTM currently activated, and (3) subset of activated LTM currently in focus of attention.[35–37] In this view, the capacity of the third level (the focus of attention) is limited to a few (typically 4) items.[36,37] In contrast, the capacity of the second level (activated LTM) is unlimited but subject to decay and interference. One of the tasks used to test this theory is the retro-cue paradigm.[38,39] In one of the variants of this task, subjects initially encode two items. Shortly after, subjects are informed by a retro-cue about which of the two items is relevant for the probe question that follows (Fig. 3A). After answering this probe question, a second retro-cue informs the subject which of the two items is relevant for the second probe question. This retro-cue could either point to the same item which was just tested in the first part or to the second item, which was unattended so far. This means that the unattended item is still relevant as it could be queried in the second part of the task. In one study by Rose and colleagues utilizing this paradigm, the two to-be-remembered items belonged to two different visual categories (e.g., text or face), thereby allowing multivariate pattern analysis to decode information about the currently maintained category.[40] This analysis revealed that decoding accuracy for the unattended category dropped sharply to chance level after the first retro-cue. Strikingly, however, information about the unattended category reappeared in the second part of the task when the second retro-cue pointed to the unattended category (Fig. 3B). This means that information about the unattended category was maintained, but this information was not decodable using this approach. This could mean that the mechanism by which attended and unattended information is maintained in WM is different (at least at the level of scalp EEG and the BOLD signal used in those experiments) or that the number of neurons carrying information in the unattended conditions is too small to be detectable using these noninvasive methods.

A second striking finding from the retro-cue experiments is the ability to transiently reactivate representations of unattended stimuli held in WM. For example, applying a brief pulse of transcranial magnetic stimulation (TMS) after the retro-cue reactivated information about the unattended category as shown by above-chance decoding.[40] This effect was observed only when TMS was applied after the first retro-cue but not after the second, when information about the unattended category was no longer relevant. Interestingly, when items from the unattended category were used as a lure during presentation of the first probe, subjects performed more errors following TMS. This result reveals that information in WM could be held in two different states: one decodable from scalp EEG or BOLD signals and the other one not decodable from these signals. Moreover, a TMS pulse could bring an unattended item back into the focus of attention. Similar results were also observed in a different retro-cue paradigm in which subjects needed to memorize the orientation of

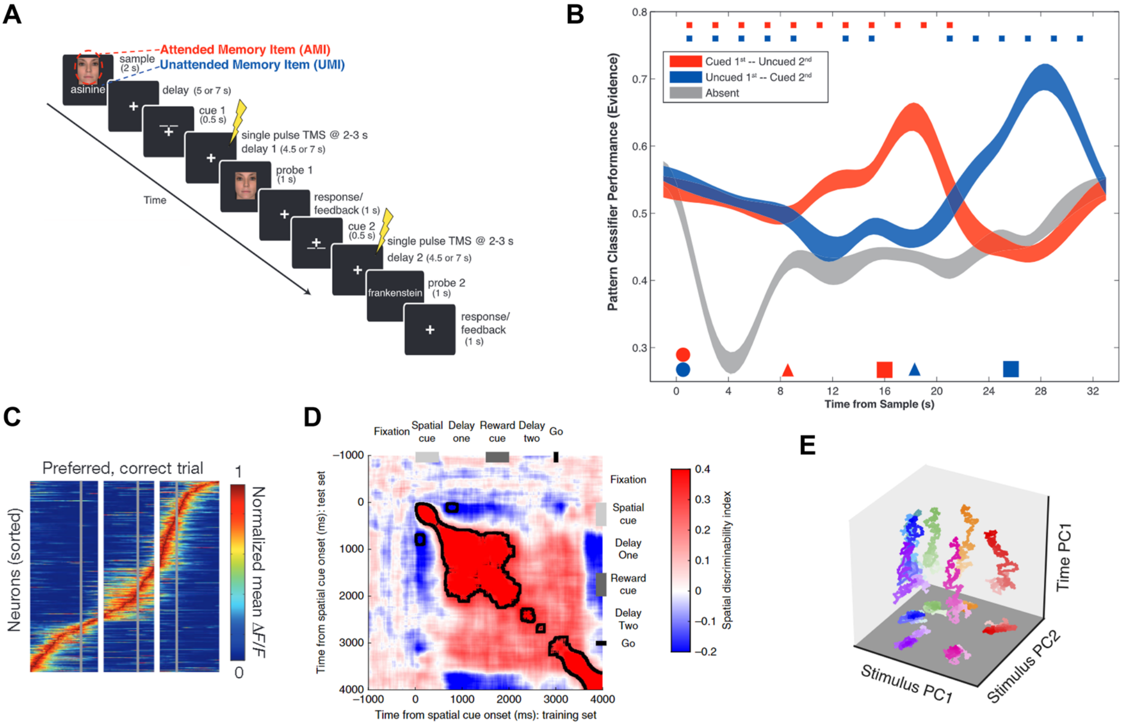

Figure 3. Different possible mechanisms of coding information in working memory. (A) Task design of retro-cue paradigm. A trial starts with presentation of two items from two different categories. After a delay period, a retro-cue indicates which of the two is relevant for the probe that follows (here: face). After a delay, a probe is shown that requires the subject to perform the task indicated by the preceding retro-cue. Afterward, a second retro-cue indicates which stimulus is relevant for the second probe, thereby leading to reactivation of items in WM. Adapted from Ref. 40. (B) Decoding of the identity of the maintained category based on the BOLD-fMRI signal in the experiment illustrated in A. Circles, triangles, and squares mark the presentation of stimuli, retro-cue, and probe, respectively. After presentation of the first retro-cue, decoding accuracy for the unattended item drops to chance but returns after the second retro-cue when subjects are instructed to bring the back into the focus of attention (blue line). Adapted from Ref. 40. (C) Dynamic coding of information in mice posterior parietal cortex measured using calcium imaging. Note the sequential activation of cells encoding spatial information. Color code indicates fluorescence intensity. Adapted from Ref. 52. (D) Mix of stable and dynamic coding in a task in which reward cues are shown as distractors during WM maintenance. Each (x, y) point shows the performance of a decoder trained and tested at a different point of time. The square-block "red" shows across-time generalization of the decoder during the first maintenance period, followed by dynamic coding (diagonal red in lower right). Adapted from Ref. 56. (E) Stable mnemonic subspace coexists with dynamic activity. Stimulus PC1 and stimulus PC2 represent the mnemonic subspace, whereas time PC1 represents the dynamic component. Colors mark locations on the screen that animals needed to memorize. Adapted from Ref. 54.

two gratings.[41] In this task, information about the orientation of gratings was not decodable from the scalp EEG signal during maintenance. However, after the researchers presented a high-contrast "ping" stimulus after the retro-cue, information about the gratings became decodable from the scalp EEG signal.

The state when information is still available to the subject but not decodable from neural activity has been referred to as an "activity-silent" state by Stokes.[42] Conceptually, we posit that activity-silent states are similar to the activated LTM level in Cowan's model. It has been suggested that information in this state is maintained by short-term synaptic plasticity (STSP) rather than by persistent firing.[28,42] Indeed, models show that a mixture of a rate code and STSP can maintain memories. For instance, Fiebig and Lansner[43] proposed a model of WM that uses a combination of rate coding and a Hebbian form of spike timing–dependent STSP. This network is able to hold multiple items in memory at the same time. In this network, well-established neuronal patterns can be held in memory in an activity-silent form for up to

8 seconds. Also, the network shows primacy and recency effects, which are commonly observed in human behavior.[44] Moreover, this network can encode and maintain novel items, which is a fundamental property of WM.

Another argument supporting the view that synaptic plasticity plays a role in WM comes from single-neuron recordings in humans.[17,18] Those studies showed that neurons in the MTL carried stimulus-specific information during WM maintenance. At first, this is a puzzling result because the MTL is principally necessary for encoding LTM, but not for WM. Indeed, in many studies, subjects with MTL lesions perform WM tasks as well as healthy controls.[45] However, under more challenging circumstances, individuals with MTL lesions also exhibit WM deficits.[46] These deficits were apparent in three situations: (1) when subjects face interference during the maintenance period, (2) when the memory load is high, and (3) when the information needs to be maintained for more than a few seconds. In all three situations, the probability that information in WM drops from the focus of attention/active maintenance is high. It is thus possible that, under this situation, WM maintenance becomes dependent on synaptic plasticity mechanisms within the MTL to recover the information that was lost from the active WM buffer. Overall, these experimental findings support a strong integration between LTM and WM, compatible with the mechanism that Cowan's model suggests. Based on this body of work, we hypothesize that only memoranda that are currently in the focus of attention are represented by PA.

A key missing piece of information is that it is currently unknown what are the mechanisms that maintain information outside of the focus of attention. This is because, so far, the retro-cue paradigm has only been used with noninvasive brain imaging methods. It is therefore possible that this information is still encoded in the firing rate of neurons, but in a manner not decodable noninvasively (by modulation of neurons in specific ways that do not result in on-average types of activity measurements). Moreover, in another study using the retrocue paradigm, it was observed that while in one area (visual cortex) information about the unattended item disappeared, other higher order cortices (IPS and FEF) continued to maintain this information.[47] This shows that in some brain areas there is no difference in the way attended and unattended information is being maintained. Moreover, in a study where naive monkeys that were never trained on a WM task passively viewed stimuli, persistent stimulus-selective activity was observed in some PFC neurons (but such activity rarely outlasted presentation of the next stimulus).[48] Together, these experiments suggest that even information outside of the focus of attention can be represented by PA. On the other hand, recordings from human MTL neurons show that stimulus-specific activity is disrupted by the onset of another stimulus if this stimulus follows the previous one rapidly.[17,18] This activity recovers after offset of the second image. This indicates that, when a subject needed to encode another stimulus, the activity representing the unattended stimulus that was already in memory was transiently disrupted. It will be critical to study how information outside of the focus of attention is being maintained at the level of single neurons and populations thereof.

Dynamics of maintenance activity

Are the same neurons representing WM content throughout the period of time an item is held in mind? A stable neuronal code is one of the main characteristics of PA.[49,50] One way to test this attribute of PA is to utilize decoders that are trained and tested at different periods of time during the maintenance period. If a decoder successfully generalizes across different time points, the subset and tuning of neurons that carry information remain stable. Alternatively, if the decoder does not generalize across time points, the subset of neurons carrying information changes as a function of time (Fig. 1). This second kind of coding has become known as *dynamic coding*, whereas the former is referred to as *static coding*.[51] For persistently active concept cells in the MTL,[17] cross-time generalization is possible across the entire task, thereby showing that human MTL neurons with PA form a stable code. However, this does not exclude the simultaneous coexistence of more dynamic forms in other groups of neurons within the same or other brain areas.

In other tasks, however, some cells with delay-period activity seem to encode WM content only at certain times after start of the maintenance period. Consider, for example, the activity of mice posterior parietal cortex neurons[52] during a run through a virtual maze. This population of neurons "tile" the delay period such that a different neuron encodes the location held in mind at different points of time (Fig. 3C). Of note, while this phenomenon has been observed in a large number of calcium imaging studies, it remains unclear what the activity of these neurons looks like at the spiking level due to the complex relationship between spiking and calcium signals.[53]

In macaques performing relatively simple paradigms, a stable code is usually observed during WM maintenance.[50] Comparing coding between encoding and maintenance, however, reveals that even in these simple paradigms a lack of cross-time generalization is sometimes apparent.[34] This is because cells can change or even invert their tuning between these two stages of the task. How then is it possible to maintain a stable representation of WM content between these two stages? Neural population analysis and theoretical modeling suggest that these two are not incompatible: highly stable mnemonic representation with robust across-time generalization can exist in the presence of dynamic coding.[34,54] While at the single-neuron level such dynamic coding presents itself as dynamic selectivity/recruitment, at the population level groups of such neurons can form perfectly stable representations that can be accessed using projection techniques to the proper mnemonic subspace[54] (Fig. 3E). Moreover, dynamic activity can also be observed in the attractor framework at transition points, that is, when stimulus-evoked activity is close to but not at the center of an attractor.[55] In this scenario, during a transient time period after stimulus removal, activity appears dynamic while the neural state settles toward the attractor center.

Interestingly, in more complex tasks, macaque PFC neurons exhibit prominent dynamic coding during the maintenance period.[34,56] For instance, this phenomenon was observed in macaque ventrolateral PFC recordings conducted by Cavanagh *et al.* during a delayed oculomotor task with additional distractors in the form of information about the reward that is presented in the middle of the maintenance period or before trial onset. One group of neurons exhibited cross-time generalization but stopped doing so if intermittent reward information was presented (Fig. 3D). A second group of neurons was characterized only by a dynamic code throughout the entire task. Also, in another paradigm by Funahashi and colleagues that required monkeys to perform an attentional and WM task at the same time, cells were also characterized by complex dynamics, with changes in tuning when comparing the attentional and WM task.[56]

The fact that this kind of dynamic activity is more frequently found in PFC during more complex tasks suggests the hypothesis that this activity might be a reflection of the part of WM that is referred to as the *central executive* in some models. The central executive is thought to be a system that controls attention and the flow of information between different memory buffers (e.g., between the phonological loop and the visuospatial sketchpad in the Baddeley model).[35,57] During complex tasks, animals need to control attention more tightly and exchange information between memory system. These functions will be used differently during the course of the task. Therefore, it is possible that dynamic coding is a reflection of the neuronal correlates of the central executive system of WM. Note that while demands on the "central executive" are higher for complex tasks, the central executive is also needed even for simple tasks that require transitions between different task phases. Therefore, dynamic activity would be expected even in simple tasks, such as delayed response tasks.[17]

Task sets and executive control of WM

Single-neuron recordings have also provided new insight into the executive control of WM.[58] For example, during a delayed oculomotor response task, animals need to keep track of the part of the task they are in (encoding, maintenance, and go). While animals require extensive experience to learn a new task set, humans can do so immediately following verbal instruction even for novel tasks. This aspect of WM is essential for flexible goal-directed behavior. This kind of flexibility is thought to rely on the PFC,[59,60] but it is poorly understood how verbal instructions are translated into task sets. One relevant result is the finding that in the human medial frontal cortex, some neurons respond selectively only in specific parts of the task, thereby demarcating transitions between task sets (encoding, maintenance, and probe).[17] These neurons did not carry information about the stimulus held in mind, but they were modulated by other aspects of the task. For example, one group of these neurons signaled the transition from encoding to maintenance, whereas another signaled the transition from maintenance to probe. Similarly, neurons in the macaque PFC have been found that respond only to the probe and whose response discriminates between whether the probe was a match or nonmatch.[61] While their specific role is unknown, this work reveals a potential substrate of task sets that might form the scaffold for the executive control of WM.

We hypothesize that task sets are necessary to properly deploy attention during WM tasks. For example, the encoding part of a WM task has different attentional demands compared to the maintenance part: during encoding, subjects need to focus their attention on the incoming stimulus, whereas during maintenance they need to change the focus of attention to internal representations of the stimuli held in WM and protect this information from outside distractors. Such control of the attentional focus is one of the functions of the central executive.[57] Note that these changes of the attentional focus occur at the time of the start of the encoding, maintenance, and probe phase of the task. Because dynamic activity is typically locked to these same periods of time (and therefore repeatable across trials), we hypothesize that dynamic activity is the mechanism that implements changes in the focus of attention.

Cellular nature of memoranda

The new findings summarized above have given rise to renewed debate regarding the different ways by which memoranda are expressed at the single-neuron level.[50,62] One the one hand, it is commonly accepted that there is by now overwhelming evidence for persistent (or delay period) activity as a principle mechanism supporting WM maintenance. On the other hand, more dynamic forms of delay-period activity have been observed in monkey PFC. For instance, research utilizing distractors or dual-task paradigms shows that static coding by PFC neurons can transiently disappear without impairing WM at the behavioral level.[56] In our opinion, these experimental findings are not incompatible with the long-held view that "persistent activity" is critical for WM maintenance. Rather, they represent the activity of different parts of the cognitive system (see conclusions).

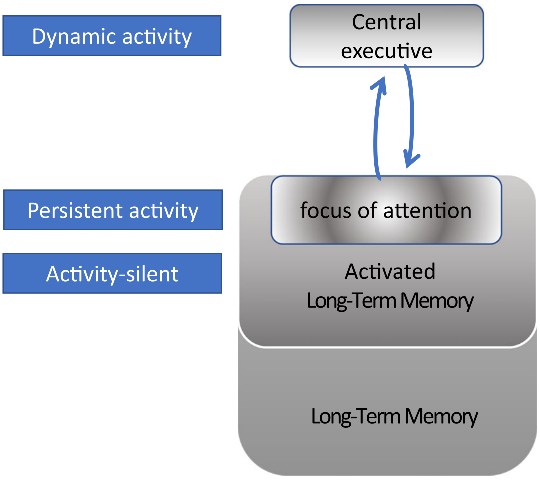

Figure 4. Illustration of Cowan's model of working memory with possible neuronal mechanism associated with each component indicated. Under this hypothesis, information currently in the focus of attention is represented by classical persistent activity. Result from the retro-cue paradigm suggests that information outside of the focus of attention is maintained by activity-silent states, a form of coding compatible with the activated long-term memory (LTM) part of the model. Finally, the central executive part of the model performs different functions as a trial progresses, thereby implementing task demands. Here, we propose that dynamic coding is a reflection of this process.

Before concluding, it is worth noting that a key difficulty in the field has been caused by an ambiguous and unclear definition of what exactly it means for a neuron to be "persistently active." A concise recent definition offered is "…we define persistent activity as memorandum-selective activity of single neurons that spans the delay interval of WM tasks."[50] While useful, there are still many ways such activity can be defined quantitatively. Among others, criteria that have been used include single-trial decodability versus on-average analysis,[63] across-time generalization, bursty versus nonbursty firing, and modulation by simultaneously recorded LFPs. For example, are neurons that increase their activity only when certain features of the LFP are also present persistently active or not?[64,65] Some have argued that such neurons are persistently active,[50] whereas others argue that they are not.[62] While it is beyond the scope of this review to advance a rigorous definition, it is important to keep in mind these discrepant definitions when relating the experimental literature to cognitive models of WM.

Table 2. Summary of the key evidence for the proposed neuronal mechanisms and their corresponding working memory concept

Concept number	Working memory concept	Neuronal mechanism	References
	Main results		
1	Part of LTM that is currently activated but which is outside the focus of attention	"Activity-silent" model	17,18,40,42
	Non-invasive imaging experiments utilizing the retro-cue paradigm show that information held in WM is decodable only if subjects focused their attention on it. Human single-neuron data shows that persistent activity is suppressed while other stimuli are being encoded.		
2	Information in focus of attention	Persistent activity	7,17,18,32,50
	Single-neuron recordings reveal stimulus-selective activity during maintenance of Working memory. The extent of drift of such persistent activity predicts memory quality. Causal intervention shows that suppression of persistent activity suppresses memoranda.		
3	Central executive	Dynamic activity	17,34,56
	Dynamic activity is reported more frequently in complex tasks, in which the central executive needs to control attention more strictly. Activity of the Central Executive changes rapidly during the course of a trial – for example, when switching from encoding to maintenance. These changes can be represented by dynamic activity. In contrast, memoranda are stable throughout the trial (persistent activity).		

NOTES: For each of the three concepts, the first row lists the working memory concept, the neuronal mechanism and key references, and the second row summarizes the main results that support this neuronal mechanism.

Conclusions

In the past decade, systems neuroscience experiments have started to provide exciting and often seemingly contradictory new insights about WM at the single-neuron level. By employing decoding, state-space modeling and optogenetics, this work has revealed unprecedented new insights into the neuronal mechanism governing WM. While these new results reveal a complex mixture of underlying mechanisms, these findings are compatible with established cognitive frameworks of WM.[35,57,58,66,67] Here, we propose that PA acts as the mechanism for representing and maintaining memoranda that are in the focus of attention (Fig. 4). This information can drop outside of the focus of attention to a less active and degraded form. Such activity-silent representations are likely supported by short-term synaptic changes. Finally, dynamic activity/coding is a good candidate to take on the role of the central executive process, because its activity will change with task demands (for instance, switching from encoding to maintenance). In contrast, mnemonic information about stimuli held in mind will stay the same. Together, this hypothesis thus suggests that the seemingly disparate experimental findings can be explained in a coherent way by relating them to different cognitive aspects of WM (see Table 2 for a summary of how the different experimental findings map onto the different aspects of WM models).

Acknowledgments

We thank all members of the Rutishauser Laboratory, Adam Mamelak, Aneta Brzezicka, and Peyman Golshani for discussion and Nand Chandravadia for help with editing. The work described here was supported by NIH (R01MH110831, U01NS103792, and P50MH094258), the McKnight Endowment for Neurosciences, and the NSF (BCS-1554105).

Competing interests

The authors declare no competing interests.

References

1. Baddeley, A. 2007. *Working Memory, Thought, and Action*. Oxford University Press.
2. Fuster, J.M. & G.E. Alexander. 1971. Neuron activity related to short-term memory. *Science* **173:** 652–654.
3. Compte, A., N. Brunel, P.S. Goldman-Rakic, *et al*. 2000. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex* **10:** 910–923.

4. Rainer, G., W.F. Asaad & E.K. Miller. 1998. Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature* **393:** 577–579.
5. Funahashi, S., C.J. Bruce & P.S. Goldman-Rakic. 1989. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* **61:** 331–349.
6. Dotson, N.M., S.J. Hoffman, B. Goodell, *et al.* 2018. Feature-based visual short-term memory is widely distributed and hierarchically organized. *Neuron* **99:** 215–226.e4.
7. Inagaki, H.K., L. Fontolan, S. Romani, *et al.* 2019. Discrete attractor dynamics underlies persistent activity in the frontal cortex. *Nature* **566:** 212–217.
8. Li, N., K. Daie, K. Svoboda, *et al.* 2016. Robust neuronal dynamics in premotor cortex during motor planning. *Nature* **532:** 459–464.
9. Chafee, M.V. & P.S. Goldman-Rakic. 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.
10. Fuster, J.M. & J.P. Jervey. 1981. Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science* **212:** 952–955.
11. Chelazzi, L., J. Duncan, E.K. Miller, *et al.* 1998. Responses of neurons in inferior temporal cortex during memory-guided visual search. *J. Neurophysiol.* **80:** 2918–2940.
12. Procyk, E. & J.P. Joseph. 2001. Characterization of serial order encoding in the monkey anterior cingulate sulcus. *Eur. J. Neurosci.* **14:** 1041–1046.
13. Isomura, Y., Y. Ito, T. Akazawa, *et al.* 2003. Neural coding of "attention for action" and "response selection" in primate anterior cingulate cortex. *J. Neurosci.* **23:** 8002–8012.
14. Watanabe, K. & S. Funahashi. 2007. Prefrontal delay-period activity reflects the decision process of a saccade direction during a free-choice ODR task. *Cereb. Cortex* **17:** i88–i100.
15. Watanabe, K. & S. Funahashi. 2014. Neural mechanisms of dual-task interference and cognitive capacity limitation in the prefrontal cortex. *Nat. Neurosci.* **17:** 601–611.
16. Romo, R., C.D. Brody, A. Hernández, *et al.* 1999. Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* **399:** 470–473.
17. Kamiński, J., S. Sullivan, J.M. Chung, *et al.* 2017. Persistently active neurons in human medial frontal and medial temporal lobe support working memory. *Nat. Neurosci.* **20:** 590–601.
18. Kornblith, S., R. Quian Quiroga, C. Koch, *et al.* 2017. Persistent single-neuron activity during working memory in the human medial temporal lobe. *Curr. Biol.* **27:** 1026–1032.
19. Minxha, J., A.N. Mamelak & U. Rutishauser. 2018. Surgical and electrophysiological techniques for single-neuron recordings in human epilepsy patients. In *Neuromethods*. W. Walz, Ed.**:** 267–293. New York, NY: Humana Press.
20. Fried, I., U. Rutishauser, M. Cerf, *et al.* 2014. *Single Neuron Studies of the Human Brain: Probing Cognition*. MIT Press.
21. Quiroga, R.Q., L. Reddy, G. Kreiman, *et al.* 2005. Invariant visual representation by single neurons in the human brain. *Nature* **435:** 1102–1107.
22. Quian Quiroga, R., A. Kraskov, C. Koch, *et al.* 2009. Explicit encoding of multimodal percepts by single neurons in the human brain. *Curr. Biol.* **19:** 1308–1313.
23. Rutishauser, U. 2019. Testing models of human declarative memory at the single-neuron level. *Trends Cogn. Sci.* **23:** 510–524.
24. Boran, E., T. Fedele, P. Klaver, *et al.* 2019. Persistent hippocampal neural firing and hippocampal–cortical coupling predict verbal working memory load. *Sci. Adv.* **5:** eaav3687.
25. Egorov, A.V., B.N. Hamam, E. Fransén, *et al.* 2002. Graded persistent activity in entorhinal cortex neurons. *Nature* **420:** 173–178.
26. Zylberberg, J. & B.W. Strowbridge. 2017. Mechanisms of persistent activity in cortical circuits: possible neural substrates for working memory. *Annu. Rev. Neurosci.* **40:** 603–627.
27. Wang, B., L. Yin, X. Zou, *et al.* 2015. A subtype of inhibitory interneuron with intrinsic persistent activity in human and monkey neocortex. *Cell Rep*. **10:** 1450–1458.
28. Wang, X.J. 2001. Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci*. **24:** 455–463.
29. Hopfield, J.J. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* **79:** 2554–2558.
30. Rutishauser, U., J.J. Slotine & R. Douglas. 2015. Computation in dynamically bounded asymmetric systems. *PLoS Comput. Biol.* **11:** e1004039.
31. Kobak, D., W. Brendel, C. Constantinidis, *et al.* 2016. Demixed principal component analysis of neural population data. *elife* **5:** 1–37.
32. Wimmer, K., D.Q. Nykamp, C. Constantinidis, *et al.* 2014. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nat. Neurosci.* **17:** 431–439.
33. Guo, Z.V., H.K. Inagaki, K. Daie, *et al.* 2017. Maintenance of persistent activity in a frontal thalamocortical loop. *Nature* **545:** 181–186.
34. Spaak, E., K. Watanabe, S. Funahashi, *et al.* 2017. Stable and dynamic coding for working memory in primate prefrontal cortex. *J. Neurosci.* **37:** 6503–6516.
35. Cowan, N. 1988. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychol. Bull.* **104:** 163–191.
36. Cowan, N. 2001. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* **24:** 87–114.
37. Cowan, N. 2005. *Working Memory Capacity*. Psychology Press.
38. Griffin, I.C. & A.C. Nobre. 2003. Orienting attention to locations in internal representations. *J. Cogn. Neurosci.* **15:** 1176–1194.
39. Landman, R., H. Spekreijse & V.A.F. Lamme. 2003. Large capacity storage of integrated objects before change blindness. *Vision Res.* **43:** 149–164.
40. Rose, N.S., J.J. LaRocque, A.C. Riggall, *et al.* 2016. Reactivation of latent working memories with transcranial magnetic stimulation. *Science* **354:** 1136–1139.
41. Wolff, M.J., J. Jochim, E.G. Akyürek, *et al.* 2017. Dynamic hidden states underlying working-memory-guided behavior. *Nat. Neurosci.* **20:** 864–871.
42. Stokes, M.G. 2015. 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn. Sci.* **19:** 394–405.

43. Fiebig, F. & A. Lansner. 2016. A spiking working memory model based on Hebbian short-term potentiation. *J. Neurosci.* **37:** 83–96.
44. Kahana, M.J. 2012. Foundations of human memory. *Theory Decis.* **61:** 368.
45. Squire, L.R., C.E.L. Stark & R.E. Clark. 2004. The medial temporal lobe. *Annu. Rev. Neurosci.* **27:** 279–306.
46. Jeneson, A. & L.R. Squire. 2012. Working memory, long-term memory, and medial temporal lobe function. *Learn. Mem.* **19:** 15–25.
47. Christophel, T.B., P. Iamshchinina, C. Yan, *et al.* 2018. Cortical specialization for attended versus unattended working memory. *Nat. Neurosci.* **21:** 494–496.
48. Meyer, T., X.L. Qi & C. Constantinidis. 2007. Persistent discharges in the prefrontal cortex of monkeys naive to working memory tasks. *Cereb. Cortex* **17**(Suppl. 1): i70–i76.
49. Kajikawa, Y. & C.E. Schroeder. 2011. How local is the local field potential? *Neuron* **72:** 847–858.
50. Constantinidis, C., S. Funahashi, D. Lee, *et al.* 2018. Persistent spiking activity underlies working memory. *J. Neurosci.* **38:** 7020–7028.
51. King, J.R. & S. Dehaene. 2014. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn. Sci.* **18:** 203–210.
52. Harvey, C.D., P. Coen & D.W. Tank. 2012. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* **484:** 62–68.
53. Jayaraman, V. 2008. Evaluating a genetically encoded optical sensor of neural activity using electrophysiology in intact adult fruit flies. *Front. Neural Circuits* **1:** 1–9.
54. Murray, J.D., A. Bernacchia, N.A. Roy, *et al.* 2016. Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proc. Natl. Acad. Sci. USA* **114:** 394–399.
55. Barbosa, J. 2018. Working memories are maintained in a stable code. *J. Neurosci.* **37:** 8309–8311.
56. Cavanagh, S.E., J.P. Towers, J.D. Wallis, *et al.* 2018. Reconciling persistent and dynamic hypotheses of working memory coding in prefrontal cortex. *Nat. Commun.* **9:** 3498.
57. Baddeley, A. 2012. Working memory: theories, models, and controversies. *Annu. Rev. Psychol.* **63:** 1–29.
58. Stokes, M.G., T.J. Buschman & E.K. Miller. 2017. Dynamic coding for flexible cognitive control. In *The Wiley Handbook of Cognitive Control*. T. Egner, Ed.: 221–241. John Wiley & Sons Ltd.
59. Miller, E.K. & J.D. Cohen. 2001. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24:** 167–202.
60. Dosenbach, N.U.F., K.M. Visscher, E.D. Palmer, *et al.* 2006. A core system for the implementation of task sets. *Neuron* **50:** 799–812.
61. Hwang, J. & L.M. Romanski. 2015. Prefrontal neuronal responses during audiovisual mnemonic processing. *J. Neurosci.* **35:** 960–971.
62. Lundqvist, M., P. Herman & E.K. Miller. 2018. Working memory: delay activity, yes! Persistent activity? Maybe not. *J. Neurosci.* **38:** 7013–7019.
63. Stokes, M. & E. Spaak. 2016. The importance of single-trial analyses in cognitive neuroscience. *Trends Cogn. Sci.* **20:** 483–486.
64. Lundqvist, M., J. Rose, P. Herman, *et al.* 2016. Gamma and beta bursts underlie working memory. *Neuron* **90:** 152–164.
65. Lundqvist, M., P. Herman, M.R. Warden, *et al.* 2018. Gamma and beta bursts during working memory readout suggest roles in its volitional control. *Nat. Commun.* **9:** 394.
66. Sreenivasan, K.K., C.E. Curtis & M. D'Esposito. 2014. Revisiting the role of persistent neural activity during working memory. *Trends Cogn. Sci.* **18:** 82–89.
67. Riley, M.R. & C. Constantinidis. 2016. Role of prefrontal persistent activity in working memory. *Front. Syst. Neurosci.* **9:** 1–14.