Routledge
Taylor & Francis Group

Theoretical distinction between functional states in working memory and their corresponding neural states

Mark G. Stokes, Paul S. Muhle-Karbe and Nicholas E. Myers

Wellcome Centre for Integrative Neuroimaging and Department of Experimental Psychology, University of Oxford, Oxford, UK

ABSTRACT

Working memory (WM) is important for guiding behaviour, but not always for the next possible action. Here we define a WM item that is currently relevant for guiding behaviour as the functionally "active" item; whereas items maintained in WM, but not immediately relevant to behaviour, are defined as functionally "latent". Traditional neurophysiological theories of WM proposed that content is maintained via persistent neural activity (e.g., stable attractors); however, more recent theories have highlighted the potential role for "activity-silent" mechanisms (e.g., short-term synaptic plasticity). Given these somewhat parallel dichotomies, functionally active and latent cognitive states of WM have been associated with storage based on persistent-activity and activity-silent neural mechanisms, respectively. However, in this article we caution against a one-to-one correspondence between functional and activity states. We argue that the principal theoretical requirement for active and latent WM is that the corresponding neural states play qualitatively different functional roles. We consider a number of candidate solutions, and conclude that the neurophysiological mechanisms for functionally active and latent WM items are theoretically independent of the distinction between persistent activity-based and activity-silent forms of WM storage.

Introduction

Information held in working memory (WM) is usually relevant for guiding future behaviour, but not necessarily the next upcoming action. Consider the following scenario: After giving a talk, an audience member asks you a two-part question. As you are preparing to answer question 1, you need to avoid distraction from question 2, lest it interfere with your first answer. However, question 2 still needs to be stored in memory so that you can answer it eventually. Ultimately, both items are of equal importance. Both need to be robustly encoded and maintained, but only the relevant one should directly influence your current behaviour. Here we define the memory item that is currently relevant as the "active" item, in the sense that it actively guides ongoing behaviour. By contrast, we define items maintained but not acted upon as "latent" items, meaning that they should not influence current processing. Latent items have the potential to be brought into an active state once the need arises, but until then are stored in a robust yet dormant format.[1]

Until recently, dominant neural models of WM required some neural activity to persist throughout the retention interval to maintain mnemonic information – with the possible exception of temporary gaps in activity that could be bridged by very short-lived phenomena, such as refractory periods (Amit & Brunel, 1997; Camperi & Wang, 1998; Wang, 2001; Wimmer et al., 2014). More recently, theorists have proposed that WM could also be maintained via "activity-silent" neural states, such as short-term synaptic plasticity (Bouchacourt & Buschman, 2019; Manohar et al., 2019; Zucker & Regehr, 2002). Although these models are not necessarily mutually exclusive, the apparent dichotomy between persistent activity and activity-silent mechanisms clearly resembles the functional distinction between active and latent cognitive states of WM (see next section). The purpose of this article is to caution against a direct correspondence between these functional and neural distinctions. We argue that the principal requirement for distinguishing functionally active from latent WM is that the neural state of active WM

should engage with ongoing processing, whereas latent WM should be functionally inert, i.e., not interfere with ongoing processes. Here we consider potential neural solutions for this and outline how they could be tested empirically.

Various neural mechanisms can support functionally latent working memory

The distinction between persistently active and activity-silent mechanisms of WM storage has generated vigorous debate (Constantinidis et al., 2018; Lundqvist et al., 2018), leading some to propose that the two neural mechanisms could serve distinct functions in the service of WM (Masse et al., 2019; Trübutschek et al., 2019). One popular proposal has been that persistent activity could be associated with attended items in WM, while other items are maintained in a more passive, activity-silent state (Kamiński & Rutishauser, 2019; LaRocque et al., 2013; LaRocque et al., 2015; Lewis-Peacock et al., 2012; Manohar et al., 2019; Olivers et al., 2011; Stokes, 2015). Although this view has intuitive appeal, the logic that functional WM states should align with this particular dichotomy of activity states has not been thoroughly evaluated in the literature.

In order to fulfil our operational definition, an active WM item should be readily available to interact with ongoing processing, whereas the latent WM item should have minimal influence. Critically, both active and latent items need to be maintained robustly. Further, latent items need to be available to be reformatted into an active state that allows them to affect behaviour when needed.

Functionally active and latent states differentially engage with ongoing processing

Behavioural studies are the main source of evidence that processing of subsequent stimuli is more reliably influenced by active WM than by latent WM. Many of these studies are based on the observation that WM can maintain templates that guide attention to memory-matching stimuli (Dowd et al., 2017; Downing, 2000; Olivers et al., 2006; Soto et al., 2005; Soto et al., 2008). This effect seems to require maintenance of WM items in an active state (Olivers et al., 2011; Van der Stigchel & Olivers, 2019), since the

attention-guiding effect rapidly subsides over multiple trial repetitions or stimulus presentations (Gayet et al., 2017; Kang & Spitzer, 2020; van Moorselaar, Theeuwes, & Olivers, 2016), presumably as the active WM item is reformatted into a non-interfering representation.

When multiple items are concurrently held in WM at different priority levels, only the active WM item appears to guide attention, while latent WM has only temporary influence (Mallett & Lewis-Peacock, 2018) or no influence at all (Greene et al., 2015; van Loon et al., 2017; van Moorselaar et al., 2014). Nevertheless, some studies have shown that multiple WM items can simultaneously guide attention (Hollingworth & Beck, 2016; Carlisle & Woodman, 2019), leaving open the possibility that, in some circumstances, latent WM items may have an unintended effect on cognition or behaviour (possibly as an effect of imperfect separation between active and latent WM via non-orthogonal coding schemes, see below).

A second line of convergent behavioural evidence comes from studies showing that novel task sets held in WM can automatically influence processing of subsequent stimuli in a secondary task performed in the maintenance delay (Liefooghe et al., 2012; Meiran et al., 2012; Muhle-Karbe et al., 2016). This effect seems to be specific to task sets held in active WM after they have been cued as relevant, while uncued task sets have no such effect (González-García et al., 2020).

By contrast, less is known about the neural basis of differential effects of active and latent WM on stimulus processing. In general, stimuli matching the contents of WM elicit larger neural responses (Awh et al., 2000; Gayet et al., 2017), consistent with increased deployment of attention. Similarly, maintaining a task set leads to motor-preparatory responses (lateralized readiness potentials) when the instructed stimuli are encountered in a secondary task, signalling that they are being processed according to the maintained stimulus-response mapping (Meiran et al., 2015), again suggesting that WM maintenance influences processing of subsequent stimuli.

At present, it remains unclear how specific these neural effects are to active WM because few studies have compared active and latent WM. A study manipulating the task-relevance of WM items found that neural markers of attention were amplified towards stimuli matching a WM item when it was task-relevant (Carlisle & Woodman, 2013), possibly because

the item was maintained in active WM. In a recent EEG study, we aimed to measure more directly how active and latent WMs influence WM-guided behaviour (Muhle-Karbe et al., 2020). Participants held two items in WM and were cued on each trial which item was active and should be compared to a probe stimulus, while the other item was latent and maintained merely for later use. Both active and latent WM items could be decoded from patterns of EEG activity, but only the quality of representation of the active item directly affected behaviour. When an item was in the active state, trial-to-trial variability in decoding predicted the efficiency of WM-based decision-making: stronger decoding of the active item led to faster performance on that trial. This was not the case for latent items, where variability in decoding did not predict behaviour on the current trial. Interestingly, however, decoding did reflect how well the item would be remembered on other trials when it was in an active state. This distinction is consistent with the idea that active and latent forms of WM storage have distinct effects on behaviour, and that latent WM minimally interferes with ongoing processing, in line with the behavioural studies cited above.

Although this study was focused on WM-based decision dynamics rather than on maintenance per se, it does highlight the key functional distinction between active and latent WM states. Below we discuss a number of neural mechanisms that could give rise to the maintenance of WM items for such functionally distinct cognitive states. First, we discuss mechanisms that segregate active and latent items in discrete neural patterns, from separation at the large-scale anatomical level, to separation of activity subspaces within the same neural population, to the frequently invoked separation via neurophysiological mechanism: persistent activity vs activity-silent states. The key property of all three mechanisms is that active and latent WM states are statistically uncorrelated (i.e., orthogonal, see also Figure 1), ensuring that active WM states can drive behaviour independently of latent WM states. Next, we discuss non-independent (i.e., non-orthogonal) coding schemes that have been proposed recently, where there is a statistical dependence between active and latent WM states, and how these might have implications for interference from latent WM on behaviour. Finally, we briefly discuss how an

alternate framework, based on separating WM representations in different phases of neural oscillations, might address the issue.

Orthogonal coding schemes for active and latent WM

If active and latent WM have different functional properties, then by definition there must be a difference in the respective neural representation. More specifically, if an item is active in one context, but latent in another context, there must be a distinct coding scheme for the same information in each context (active vs latent; see Figure 1). If they were represented in exactly the same neural state, they would have exactly the same functional properties. Further, we suggest that the difference between active and latent WM states should be *qualitative*, not just *quantitative*. Both active and latent memories need to be maintained robustly: the key difference in not the strength of the coding, but its functional properties. Finally, if active and latent neural states are maintained by independent, or orthogonal, coding schemes, interference should be minimized.

Anatomical separation

Perhaps the simplest means by which the brain could maintain active and latent WM items independently is via storage in distinct brain areas (Figure 1A, left). Consistent with this idea, a recent fMRI study found that active items could be decoded from BOLD activity patterns in a distributed network comprising early visual, parietal, and prefrontal areas. By contrast, latent items could be decoded from activity patterns in the intraparietal sulcus and the frontal eye fields (Christophel et al., 2018), suggestive of a division of labour between brain areas in coding for active and latent WM items. In principle, these findings meet our criterion for qualitatively different coding schemes (Figure 1A, middle and right), since behaviour could be selectively driven by those brain areas representing only the active item. Testing how decoding strength in different brain areas relates to the quality of WM-based decisions will be central to evaluating the anatomical separation hypothesis. Ultimately, causal interventions (e.g., Daie et al., 2019) may help determine whether only regions exclusively coding for active WM drive behaviour.
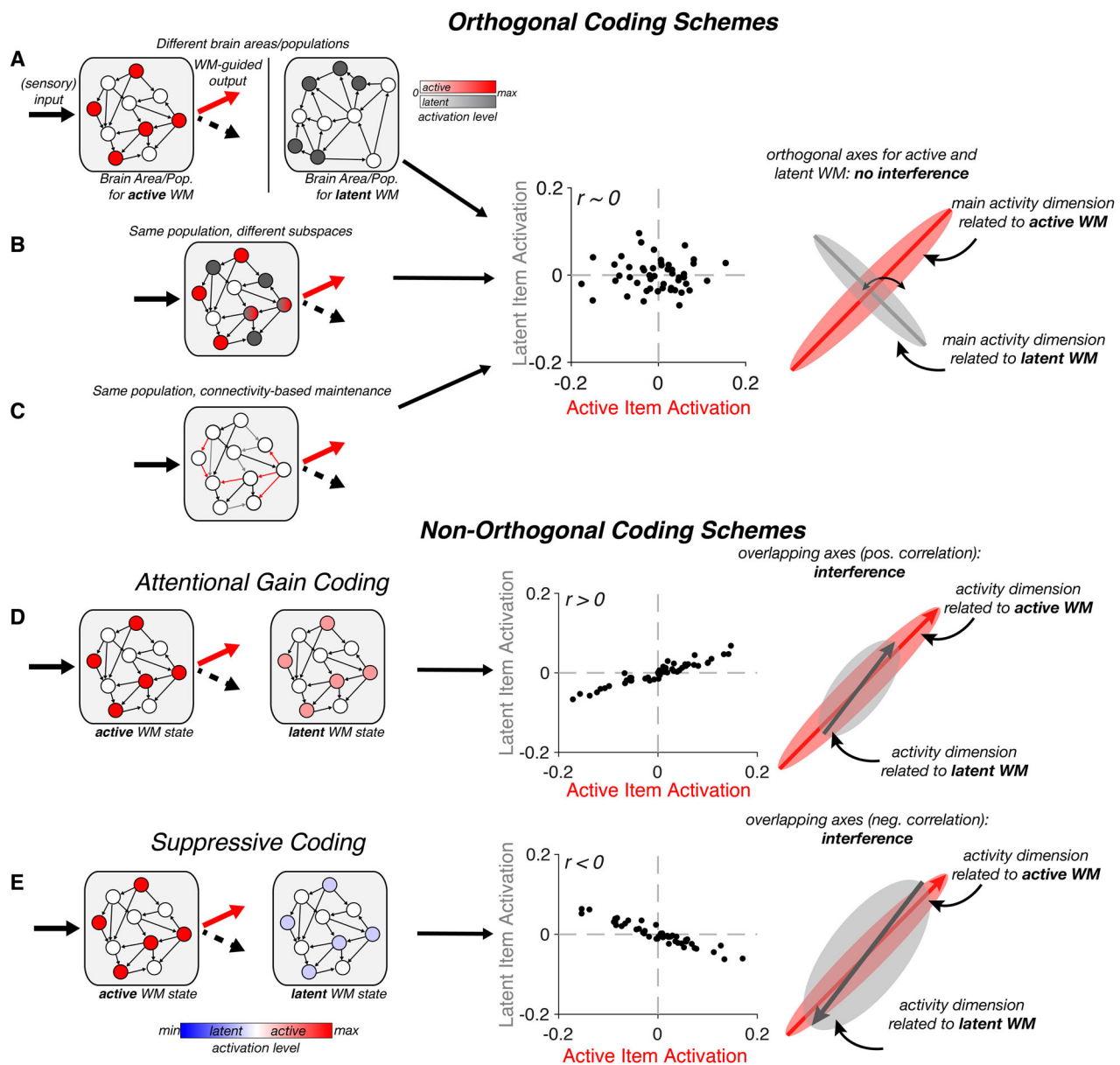
Orthogonal Coding Schemes

Non-Orthogonal Coding Schemes

Figure 1. Summary of possible coding schemes for active vs. latent WM. Rows show different putative coding mechanisms for active versus latent WM. Left-hand column: Circuit-level depiction of various coding schemes in an example neural population. Each grey square represents a WM-coding neural population. Within the population, circles represent coding units (neurons), and arrows represent directed connections. Activated units are shown in colour (active: red, latent: grey or blue). Middle column: Correlation between activation patterns for items in an active (x-axis) or a latent (y-axis) state. Individual points indicate units. Correlations are exaggerated for illustration. Right-hand column: Neural state-space representation. When reduced to their most informative dimensions, neural patterns for active or latent items may occupy different subspaces. The extent of their overlap is a reflection of how correlated patterns are for active and latent WM. (A-C). Various coding schemes leading to orthogonal representations (no correlation between active and latent patterns). (A). Separate brain areas or separable neural populations. (B). Separate patterns in the same neural population. (C). Connectivity-based (i.e., activity-silent storage) can also separate active from latent WM by changing the weights of different connections in the population. (D-E). Non-orthogonal coding schemes. (D). Attention Gain coding separates active from latent WM through differences in amplitude, rather than different patterns. (E) Similarly, suppressive coding stores latent WM in the same neural pattern, but through a reversal of activity, leading to anti-correlated activity patterns that nevertheless occupy the same neural subspace.

It is also important to note that fMRI findings of decodable delay signals do not strictly imply that WM items are encoded in persistent activity, rather than in "activity silent" states. Both could in principle result in decodable patterns in the BOLD response: Persistent activity should drive statistically separable

patterns across voxels, but it is also possible that activity-silent states can be detected in separable patterns of spontaneous activity (Sugase-Miyamoto et al., 2008). Given the indirect coupling of spiking activity and BOLD (e.g., Logothetis et al., 2001), there are probably even more indirect possibilities that complicate the correspondence between decodable BOLD signals and the underlying neurophysiological mechanism. Nevertheless, whatever the underlying activity state, maintenance in anatomically segregated brain areas fulfils our basic orthogonality requirement.

Separation by different subpopulations (or subspaces) within brain areas

A second possibility is that active and latent WM items are stored within the same brain area(s), or even in an overlapping neural population, without causing interference (Figure 1B). The main prerequisite for this coding scheme is that item-specific activity patterns for the active item are uncorrelated with item-specific patterns when the same item is latent (i.e., active and latent states are statistically separable). This may correspond to latent item patterns occupying the null space of the optimal readout weights for the active item. While the theoretical appeal of such a coding scheme has been noted (Druckmann & Chklovskii, 2012; Spaak et al., 2017), supporting evidence is still relatively scarce, and human evidence is still lacking. It is worth noting that such a scheme would imply that some cells in the population exhibit nonlinear mixed selectivity (Fusi et al., 2016) since their response to a particular WM item would depend on its functional state. The involvement of mixed selectivity in WM maintenance has been demonstrated in other contexts (Parthasarathy et al., 2017; Rigotti et al., 2013), where mixed selectivity neurons have been shown to be of particular relevance to behaviour, highlighting the potential utility of context-dependent codes.

A recent WM study recording neurons in prefrontal, parietal, and visual cortex (Panichello & Buschman, 2020) found that cueing a WM item as relevant led to a transformation of activity into a new subspace that was orthogonal to the subspace coding for the same item prior to the cue. This result can be interpreted as evidence that orthogonalisation plays a role in distinguishing active from latent WM. In a related study (Yoo & Hayden, 2020) recording from neurons in the orbitofrontal and ventromedial prefrontal cortex, two stimuli that were both needed for a reward-guided decision were maintained across a delay period in orthogonal subspaces. This separation within the same neural populations could allow downstream brain areas to be driven entirely by one stimulus without interference from the other. A similar mechanism has been demonstrated in movement planning, where premotor cortex maintains a planned movement in a latent state that is "invisible" to the motor cortex until it needs to be executed (Elsayed et al., 2016; Kaufman et al., 2014). Notably, as for the anatomical coding scheme outlined above, such orthogonal patterns could be maintained both via persistent activity, or through activity-independent means (e.g., Hopfield, 1982).

Separation by neurophysiological mechanism

Finally, we consider the proposal that functionally active states are supported by elevated neural activity, whereas functionally latent states correspond to activity-silent mnemonic mechanisms. This distinction has been suggested by a number of authors (e.g., Kamiński & Rutishauser, 2019; LaRocque et al., 2013, 2015; Lewis-Peacock et al., 2012; Manohar et al., 2019; Olivers et al., 2011; Stokes, 2015). In the framework developed here, this division of labour is only helpful if it confers differential functional properties on active and latent WMs. As highlighted above, it does not bear on the basic maintenance demands: both active and latent memories need to be maintained robustly. Nevertheless, a division of labour between different candidate neurophysiological mechanisms (a stable attractor state based on persistent activity, or short-term synaptic plasticity) could satisfy our key requirement for orthogonal representation (Figure 1C). However, it is also often further assumed that the neurophysiological dichotomy between persistent activity and activity-silent maintenance naturally aligns with the functional dichotomy between active and latent WM. The intuition seems to be that elevated neural activity uniquely influences WM-guided behaviour (e.g., via changes in state-dependent processing), and therefore is better suited to active WM, whereas activity-silent mechanisms are effectively functionally dormant (see, e.g., Lewis-Peacock et al., 2012). However, it is important to point out that activity-silent mechanisms

are not inherently functionally dormant. On the contrary, temporary changes in synaptic connectivity can have a direct functional impact on subsequent processing. For instance, encoding an active item via altered responsivity in the relevant network could allow subsequent input (i.e., a memory probe) to evoke activity that will generate an appropriate response (Stokes et al., 2013), without requiring sustained activity (Manohar et al., 2019; Mongillo et al., 2008). Therefore, both forms of maintenance (persistent activity, e.g., Mante et al., 2013; Remington et al., 2018, and activity-silent connectivity patterns, e.g., Bouchacourt & Buschman, 2019; Manohar et al., 2019) can guide decision-making and support active WM states. Similarly, as we have described above, functionally latent representations could also be maintained via persistent activity as long as they are qualitatively distinct from the corresponding active representation.

It is also important to note that other factors likely determine the extent to which a WM item is associated with persistent activity or activity-silent states. For example, it has recently been proposed that elevated activity could be a signature of current processing or transformation of WM items, rather than storage per se (Masse et al., 2020). If such transformations are more likely to occur on active WM items (as has been proposed previously, see Lewis-Peacock et al., 2012), this could explain common findings such as active WM being decoded from BOLD activity, while latent items are not.

Non-orthogonal coding schemes

A number of alternative proposals for neural differences between active and latent items fall into the general category of non-orthogonal coding (Figure 1D-E). For example, Schneegans and Bays (2017) argued that active and latent items are encoded in the same neural patterns and differ only quantitatively in their level of activation (see also Chun, 2011; Kiyonaga & Egner, 2013). They describe an attractor model with an attention parameter that modulates the gain of activity coding for items cued as relevant, compared to latent items (Figure 1D, left). Importantly, persistent activity of items prior to cue presentation is just attenuated, not abolished. If the latent item becomes relevant, activity is increased to the activation level of an active item, allowing more accurate

readout. It is worth noting that this model implies that basic maintenance only requires a relatively low energy persistent activity state, whereas the additional activity for attended/active items serves a distinct purpose (e.g., to allow for more reliable readout of the attended item).

In contrast to the orthogonal coding schemes listed above, a difference in activity level between active and latent items means that the underlying patterns are positively correlated (Figure 1D, middle), which presumably could lead to greater cross-talk between active and latent states (Figure 1D, right). For example, the latent item could distort the readout population's estimate of the active item, or create confusion between active and latent items. The severity of this confusion should depend on the relative activation strength of the latent item. One consequence of this should be a trade-off between confusability with the latent item, and more general durability of the memory. While this trade-off could help explain classic WM capacity limits (Ma et al., 2014), there is some behavioural evidence that formerly latent items can be restored to an active status with little information loss (Hollingworth & Maxcey-Richard, 2013; Rerko & Oberauer, 2013; also Oberauer, 2002; and Nee & Jonides, 2014; but see, e.g., Rerko et al., 2014, for costs to latent WM).

Another non-orthogonal coding scheme for active and latent items is suppressive coding. Suppressive coding has been recognized as a general feature of WM delay activity. This manifests as a reversal of selectivity of a proportion of neurons (or of activation patterns measured with BOLD) during a memory delay, relative to encoding. Reversed selectivity is consistent with the idea that some memory-selective cells reduce their firing rate below baseline during WM maintenance. This effect has been demonstrated in sensory areas (Linke et al., 2011; Lorenc et al., 2020; Zaksas & Pasternak, 2006), prefrontal cortex (Fuster, 1973; Hussar & Pasternak, 2012; Lara & Wallis, 2014; Zhou et al., 2012), and parietal cortex (Zhou et al., 2012). Building on activity reversals relative to stimulus encoding, two recent fMRI studies have suggested that a WM item can also reverse its activity profile between active and latent states (van Loon et al., 2018; Yu et al., 2020). The studies measured BOLD signals in visual and parietal cortex to decode active and latent WM items, and found that when a classifier was trained to discriminate the active WM and was

then applied to identify the latent item, its performance dropped below chance level. In other words, item-specific patterns are anti-correlated between the active and latent state. One mechanism to achieve this could be suppression of item-selective neurons when that item enters a latent state (Figure 1E, left and middle).

A suppression or reversal of coding for latent items might seem like a means of reducing their influence on behaviour. However, even a negative relationship means that a meaningful portion of the activity related to the latent item falls into the subspace coding for the active item (Figure 1E, right). This negative correlation potentially suffers from the same problem as any mechanism relying on positively correlated patterns between active and latent states: readout trained to discriminate the active item might be influenced by the identity of the latent item. In particular, suppressive coding of the latent item should drive readout of the active item to be *less* similar to the latent one than it actually was, possibly leading to mnemonic repulsion between active and latent items. Interestingly, such repulsion has been reported in certain cases at the behavioural level (e.g., Myers et al., 2018; see also Almeida et al., 2015; Nassar et al., 2018). This points to the possibility that suppressive coding could be adaptive in some task environments when similar items need to be disambiguated (e.g., Geng et al., 2017). More generally, non-orthogonal coding mechanisms could explain why behaviour can be temporarily guided by latent WM (Mallett & Lewis-Peacock, 2018).

Temporal separation via oscillations

An alternate form of separating active and latent WM contents is via temporal coding, for instance through periodic reactivation of individual WM items at different phases of an ongoing slow oscillation (Axmacher et al., 2010; Bahramisharif et al., 2018; Jensen & Lisman, 1996; Lisman & Idiart, 1995). To our knowledge, phase separation of active and latent WM items has not yet been demonstrated, but this appears to be a plausible mechanism. For instance, inter-areal phase synchronization (Johnson et al., 2018; Spellman et al., 2015) could ensure that a downstream population is driven only by item reactivations at the optimal phase (reserved for the active item),

while latent items reactivate at suboptimal phases, ensuring minimal interference with subsequent processing. This would achieve a similar aim as segregating activity into distinct subspaces.

Transforming latent into active WM

Our aim is not to advocate for any specific means of distinguishing active from latent WM. However, it is worth considering how plausibly the possibilities outlined above could fulfil one key demand: transferring WM items from a latent to an active state (or vice versa). The separation by neurophysiological mechanism (activity vs activity-silent storage) seems to provide the most straightforward solution: a latent WM item can be moved to an active state simply by activating the population storing the latent item in its synaptic weights. Several computational models have shown that pattern completion can lead a population into an attractor state coding for the now active item (Manohar et al., 2019; Oberauer & Lin, 2017). Attentional gain coding and suppressive coding could also allow for relatively straightforward transformations, by either amplifying existing activity (attentional gain) or rebounding from a suppression of activity. By contrast, anatomical separation requires an apparently more complex transfer of information from one brain area (coding for latent WM) to another (coding for active WM). While this has not been demonstrated for WM, it has been shown in other contexts (e.g., Crowe et al., 2013). The study by Christophel et al. (2018) also strongly suggests this possibility, since their paradigm required maintaining two items until a retrospective cue indicated which to prioritize, presumably triggering its reactivation in visual cortex. Similarly, separation by activity subspace within a brain area faces the challenge of shifting neural activity from one subspace to another. While this has not been demonstrated specifically for transitions of latent to active WM, it has been observed in a number of other contexts (e.g., Tang et al., 2019), and particular in transitioning representations from a neutral to a prioritized state after a retrospective cue (Panichello & Buschman, 2020). Since the mechanism for moving information across brain areas or subspaces is unknown, it cannot be ruled out that such a transformation could reintroduce interference between active and latent WM that had been avoided by orthogonalising them in the first place.

Future directions

We have laid out a variety of theoretical mechanisms for the storage of functionally active or latent items in WM. Since the key constraint for latent items is that they should not interfere with current behaviour or storage of the active item, the main constraint on possible storage mechanisms is not whether it is persistently active or activity-silent, but rather the orthogonality of the respective coding schemes. Although separation by neurophysiological mechanisms (persistent activity vs activity-silent maintenance) could fulfil this key constraint, it is only one of a larger set of possible solutions. Importantly, it is possible that active and latent items could both be maintained using the same kind of activity state (persistent activity or activity silent maintenance), so long as the mnemonic states are orthogonal: e.g., separate brain areas, overlapping but distinct neural populations, or non-overlapping activity subspaces of the same population.

The distinction between active and latent items echoes the distinction between attentional templates and accessory memory items made by Olivers et al. (2011). The authors distinguished between different neural mechanisms that could underlie the storage of accessory memory items so that it does not drive attentional capture. Our framework can be thought of as an extension of this idea. Attentional capture by stimuli matching the active item (but not the latent item) is one means by which the active item may influence cognition or behaviour. In this instance, the maintenance of the active item would be expected to exert top-down influence on sensory areas so that they preferentially process matching input. In a more general framework, this can be interpreted as one of several possible downstream consequences of the active item that needs to be avoided by the latent item. As we showed recently (Muhle-Karbe et al., 2020), the same principle should also extend to WM-based decision-making (see also Myers et al., 2015).

Identifying which of the outlined mechanisms support the distinction between active and latent WM requires robust methods for the identification of neural coding mechanisms. However, significant challenges remain for testing candidate neurophysiological coding schemes such as persistent activity and activity-silent coding. In particular, activity-silent states are fundamentally challenging to infer, given that most methods in neuroscience measure some form (or correlate) of neural activity. Recently, we developed an impulse-response approach to "ping" activity-silent neural states by measuring the brain's response to task-irrelevant driving input, providing a theoretical potential to track the behaviour of a greater variety of mnemonic states (Wolff et al., 2015; Wolff et al., 2017; see also Rose et al., 2016). While this approach will be useful for enhancing our sensitivity for detecting memory signals that are otherwise undetectable (for whatever reason), on its own it does not strictly adjudicate between the alternative neurophysiological mechanisms. For instance, Schneegans and Bays (2017) have illustrated how low levels of persistent activity could underlie an apparently activity-silent code. Nevertheless, this approach can provide a useful complement to more standard measures of delay-period activity because it more sensitively detects activity related to latent WM items (Wolff et al., 2017). This, in turn, could be used to test more in-depth questions, such as whether active and latent items share the same code.

Definitive evidence for activity-silent mechanisms will ultimately require specific evidence of the supposed underlying processes, such as temporary connectivity changes (e.g., periodic refreshing or reactivation of a memory representation, Mongillo et al., 2008), or intrinsic gain modulation (e.g., Stroud et al., 2018). Moreover, the ability to identify how the brain separates WM representations will always be limited by what current methods can detect. Similarly, even when items appear to share the same coding mechanism, we cannot rule out undiscovered differences in coding that are invisible to current methods. At the same time, inferring persistent, uninterrupted activity is not trivial either. Elevated firing during delay periods could reflect transient non-maintenance processes, which can appear to be persistent firing when averaged over many trials (Miller et al., 2018).

More generally, it may be insufficient to rely on decodability alone to infer a putative mnemonic state. Decodability has become a ubiquitous marker of WM maintenance (for an overview on WM and decoding, see Christophel et al., 2017), but we propose that future work will need to focus less on the mere presence or absence of decodable neural patterns, and more on the functional properties of

candidate neural states. This is important because a decodable neural state could be epiphenomenal to WM (e.g., reflect mental imagery or probe expectation, rather than maintenance per se, see Stokes, 2011). Moreover, even if the neural state is necessary for WM, it is still critical to understand *how* it influences neural response dynamics to gain a mechanistic understanding of the underlying process (rather than simply identifying the brain area maintaining the WM engram).

Conclusions

In conclusion, we caution against a direct equivalence between functional states in working memory and their corresponding neural states. The key theoretical constraint is that active and latent WMs should be maintained via qualitatively distinct neural states. Within these theoretical constraints, the precise mechanisms of maintenance for either type of WM remains an empirical question. There remain major challenges associated with establishing the neurophysiological mechanisms of maintenance. We argue that focusing on the functional behaviour of putative mnemonic states will be an important future direction.

Note

1. Note that related distinctions between priority states in WM have previously been made. Classic WM models distinguish between attended and unattended items (Cowan, 2001; Oberauer, 2002), between items that guide attention and those that do not (Olivers et al., 2011), or between prioritized and (temporarily) unprioritized items (Lewis-Peacock et al., 2012). In the present article, we focus on the functional differences between representational states, i.e., how they differentially affect subsequent processing (see also Myers et al., 2017). Therefore, here we prefer to use the terms "active" and "latent" to emphasize functional consequences and to avoid connoting a particular role for attention in the process.

Acknowledgments

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

ORCID

Nicholas E. Myers http://orcid.org/0000-0001-5599-3044

References

Almeida, R., Barbosa, J., & Compte, A. (2015). Neural circuit basis of visuo-spatial working memory precision: A computational and behavioral study. *Journal of Neurophysiology*, *114*(3), 1806–1818. https://doi.org/10.1152/jn.00362.2015

Amit, D. J., & Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex*, *7*(3), 237–252. https://doi.org/10.1093/cercor/7.3.237

Awh, E., Anllo-Vento, L., & Hillyard, S. A. (2000). The role of spatial selective attention in working memory for locations: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, *12*(5), 840–847. https://doi.org/10.1162/089892900562444

Axmacher, N., Henseler, M. M., Jensen, O., Weinreich, I., Elger, C. E., & Fell, J. (2010). Cross-frequency coupling supports multi-item working memory in the human hippocampus. *Proceedings of the National Academy of Sciences*, *107*(7), 3228–3233. https://doi.org/10.1073/pnas.0911531107

Bahramisharif, A., Jensen, O., Jacobs, J., Lisman, J., & Tong, F. (2018). Serial representation of items during working memory maintenance at letter-selective cortical sites. *PLoS Biology*, *16*(8), e2003805. https://doi.org/10.1371/journal.pbio.2003805

Bouchacourt, F., & Buschman, T. J. (2019). A flexible model of working memory. *Neuron*, *103*(1), 147–160.e8. https://doi.org/10.1016/j.neuron.2019.04.020

Camperi, M., & Wang, X. J. (1998). A model of visuospatial working memory in prefrontal cortex: Recurrent network and cellular bistability. *Journal of Computational Neuroscience*, *5*(4), 383–405. https://doi.org/10.1023/A:1008837311948

Carlisle, N. B., & Woodman, G. F. (2013). Reconciling conflicting electrophysiological findings on the guidance of attention by working memory. *Attention, Perception, & Psychophysics*, *75*(7), 1330–1335. https://doi.org/10.3758/s13414-013-0529-7

Carlisle, N. B., & Woodman, G. F. (2019). Quantifying the attentional impact of working memory matching targets and

distractors. *Visual Cognition*, *27*(5-8), 452–466. https://doi.org/10.1080/13506285.2019.1634172

Christophel, T. B., Iamshchinina, P., Yan, C., Allefeld, C., & Haynes, J.-D. (2018). Cortical specialization for attended versus unattended working memory. *Nature Neuroscience*, 1–11. https://doi.org/10.1038/s41593-018-0094-4

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J.-D. (2017). The distributed nature of working memory. *Trends in Cognitive Sciences*, *21*(2), 111–124. https://doi.org/10.1016/j.tics.2016.12.007

Chun, M. M. (2011). Visual working memory as visual attention sustained internally over time. *Neuropsychologia*, *49*(6), 1407–1409. https://doi.org/10.1016/j.neuropsychologia.2011.01.029

Constantinidis, C., Funahashi, S., Lee, D., Murray, J. D., Qi, X.-L., Wang, M., & Arnsten, A. F. T. (2018). Persistent spiking activity underlies working memory. *Journal of Neuroscience*, *38*(32), 7020–7028. https://doi.org/10.1523/JNEUROSCI.2486-17.2018

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Crowe, D. A., Goodwin, S. J., Blackman, R. K., Sakellaridi, S., Sponheim, S. R., MacDonald, A. W., & Chafee, M. V. (2013). Prefrontal neurons transmit signals to parietal neurons that reflect executive control of cognition. *Nature Neuroscience*, *16*(10), 1484–1491. https://doi.org/10.1038/nn.3509

Daie, K., Svoboda, K., & Druckmann, S. (2019). Targeted photo-stimulation uncovers circuit motifs supporting short-term memory. *bioRxiv*, *12*, 623785. https://doi.org/10.1101/623785

Dowd, E. W., Pearson, J. M., & Egner, T. (2017). Decoding working memory content from attentional biases. *Psychonomic Bulletin Review*, *24*(4), 1252–1260. https://doi.org/10.3758/s13423-016-1204-5

Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science*, *11*(6), 467–473. https://doi.org/10.1111/1467-9280.00290

Druckmann, S., & Chklovskii, D. B. (2012). Neuronal circuits underlying persistent representations despite time varying activity. *Current Biology*, *22*(22), 2095–2103. https://doi.org/10.1016/j.cub.2012.08.058

Elsayed, G. F., Lara, A. H., Kaufman, M. T., Churchland, M. M., & Cunningham, J. P. (2016). Reorganization between preparatory and movement population responses in motor cortex. *Nature Communications*, *7*(1), 13239–13215. https://doi.org/10.1038/ncomms13239

Fusi, S., Miller, E. K., & Rigotti, M. (2016). Why neurons mix: High dimensionality for higher cognition. *Current Opinion in Neurobiology*, *37*, 66–74. https://doi.org/10.1016/j.conb.2016.01.010

Fuster, J. M. (1973). Unit activity in prefrontal cortex during delayed-response performance: Neuronal correlates of transient memory. *Journal of Neurophysiology*, *36*(1), 61–78. https://doi.org/10.1152/jn.1973.36.1.61

Gayet, S., Guggenmos, M., Christophel, T. B., Haynes, J.-D., Paffen, C. L. E., Van der Stigchel, S., & Sterzer, P. (2017).

Visual working memory enhances the neural response to matching visual input. *Journal of Neuroscience*, *37*(28), 6638–6647. https://doi.org/10.1523/JNEUROSCI.3418-16.2017

Geng, J. J., DiQuattro, N. E., & Helm, J. (2017). Distractor probability changes the shape of the attentional template. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(12), 1993–2007. https://doi.org/10.1037/xhp0000430

González-García, C., Formica, S., Liefooghe, B., & Brass, M. (2020). Attentional prioritization reconfigures novel instructions into action-oriented task sets. *Cognition*, *194*, 104059. https://doi.org/10.1016/j.cognition.2019.104059

Greene, C. M., Kennedy, K., & Soto, D. (2015). Dynamic states in working memory modulate guidance of visual attention: Evidence from an *n*-back paradigm. *Visual Cognition*, *23*(5), 546–560. https://doi.org/10.1080/13506285.2015.1052033

Hollingworth, A., & Beck, V. M. (2016). Memory-based attention capture when multiple items are maintained in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(7), 911–917. https://doi.org/10.1037/xhp0000230

Hollingworth, A., & Maxcey-Richard, A. M. (2013). Selective maintenance in visual working memory does not require sustained visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(4), 1047–1058. https://doi.org/10.1037/a0030238

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, *79*(8), 2554–2558. https://doi.org/10.1073/pnas.79.8.2554

Hussar, C. R., & Pasternak, T. (2012). Memory-guided sensory comparisons in the prefrontal cortex: Contribution of putative pyramidal cells and interneurons. *Journal of Neuroscience*, *32*(8), 2747–2761. https://doi.org/10.1523/JNEUROSCI.5135-11.2012

Jensen, O., & Lisman, J. E. (1996). Novel lists of 7 +/− 2 known items can be reliably stored in an oscillatory short-term memory network: Interaction with long-term memory. *Learning & Memory*, *3*(2-3), 257–263. http://doi.org/10.1101/lm.3.2-3.257

Johnson, E. L., Adams, J. N., Solbakk, A.-K., Endestad, T., Larsson, P. G., Ivanovic, J., Meling, T. R., Lin, J. J., Knight, R. T., & Rushworth, M. (2018). Dynamic frontotemporal systems process space and time in working memory. *PLoS Biology*, *16*(3), e2004274. https://doi.org/10.1371/journal.pbio.2004274

Kamiński, J., & Rutishauser, U. (2019). Between persistently active and activity-silent frameworks: Novel vistas on the cellular basis of working memory. *Annals of the New York Academy of Sciences*, *61*, 267. https://doi.org/10.1111/nyas.14213

Kang, Z., & Spitzer, B. (2020). Concurrent visual working memory bias in perception and decision making. *bioRxiv*, *26*, 2020.07.16.191445. https://doi.org/10.1101/2020.07.16.191445

Kaufman, M. T., Churchland, M. M., Ryu, S. I., & Shenoy, K. V. (2014). Cortical activity in the null space: Permitting preparation without movement. *Nature Neuroscience*, 17(3), 440–448. https://doi.org/10.1038/nn.3643

Kiyonaga, A., & Egner, T. (2013). Working memory as internal attention: Toward an integrative account of internal and external selection processes. *Psychonomic Bulletin & Review*, 20(2), 228–242. https://doi.org/10.3758/s13423-012-0359-y

Lara, A. H., & Wallis, J. D. (2014). Executive control processes underlying multi-item working memory. *Nature Neuroscience*, 17(6), 876–883. https://doi.org/10.1038/nn.3702

LaRocque, J. J., Eichenbaum, A. S., Starrett, M. J., Rose, N. S., Emrich, S. M., & Postle, B. R. (2015). The short- and long-term fates of memory items retained outside the focus of attention. *Memory & Cognition*, 43(3), 453–468. https://doi.org/10.3758/s13421-014-0486-y

LaRocque, J. J., Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2013). Decoding attended information in short-term memory: An EEG study. *Journal of Cognitive Neuroscience*, 25(1), 127–142. https://doi.org/10.1162/jocn_a_00305

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural evidence for a distinction between short-term memory and the focus of attention. *Journal of Cognitive Neuroscience*, 24(1), 61–79. https://doi.org/10.1162/jocn_a_00140

Liefooghe, B., Wenke, D., & De Houwer, J. (2012). Instruction-based task-rule congruency effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(5), 1325–1335. https://doi.org/10.1037/a0028148

Linke, A. C., Vicente-Grabovetsky, A., & Cusack, R. (2011). Stimulus-specific suppression preserves information in auditory short-term memory. *Proceedings of the National Academy of Sciences*, 108(31), 12961–12966. https://doi.org/10.1073/pnas.1102118108

Lisman, J. E., & Idiart, M. (1995). Storage of 7 +/− 2 short-term memories in oscillatory subcycles. *Science*, 267(5203), 1512–1515. https://doi.org/10.1126/science.7878473

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843), 150–157. https://doi.org/10.1038/35084005

Lorenc, E. S., Vandenbroucke, A. R. E., Nee, D. E., de Lange, F. P., & D'Esposito, M. (2020). Dissociable neural mechanisms underlie currently-relevant, future-relevant, and discarded working memory representations. *Scientific Reports*, 10(1), 11195–11117. https://doi.org/10.1038/s41598-020-67634-x

Lundqvist, M., Herman, P., & Miller, E. K. (2018). Working memory: Delay activity, yes! Persistent activity? Maybe not. *Journal of Neuroscience*, 38(32), 7013–7019. https://doi.org/10.1523/JNEUROSCI.2485-17.2018

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, 17(3), 347–356. https://doi.org/10.1038/nn.3655

Mallett, R., & Lewis-Peacock, J. A. (2018). Behavioral decoding of working memory items inside and outside the focus of attention. *Annals of the New York Academy of Sciences*, 1424(1), 256–267. https://doi.org/10.1111/nyas.13647

Manohar, S. G., Zokaei, N., Fallon, S. J., Vogels, T. P., & Husain, M. (2019). Neural mechanisms of attending to items in working memory. *Neuroscience & Biobehavioral Reviews*, 101, 1–12. https://doi.org/10.1016/j.neubiorev.2019.03.017

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503(7474), 78–84. https://doi.org/10.1038/nature12742

Masse, N. Y., Rosen, M. C., & Freedman, D. J. (2020). Reevaluating the role of persistent neural activity in short-term memory. *Trends in Cognitive Sciences*, 24(3), 242–258. https://doi.org/10.1016/j.tics.2019.12.014

Masse, N. Y., Yang, G. R., Song, H. F., Wang, X.-J., & Freedman, D. J. (2019). Circuit mechanisms for the maintenance and manipulation of information in working memory. *Nature Neuroscience*, 22(7), 1159–1167. https://doi.org/10.1038/s41593-019-0414-3

Meiran, N., Cole, M. W., & Braver, T. S. (2012). When planning results in loss of control: Intention-based reflexivity and working-memory. *Frontiers in Human Neuroscience*, 6, 104. https://doi.org/10.3389/fnhum.2012.00104

Meiran, N., Pereg, M., Kessler, Y., Cole, M. W., & Braver, T. S. (2015). Reflexive activation of newly instructed stimulus–response rules: Evidence from lateralized readiness potentials in no-go trials. *Cognitive, Affective, & Behavioral Neuroscience*, 15(2), 365–373. https://doi.org/10.3758/s13415-014-0321-8

Miller, E. K., Lundqvist, M., & Bastos, A. M. (2018). Working memory 2.0. *Neuron*, 100(2), 463–475. https://doi.org/10.1016/j.neuron.2018.09.023

Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic theory of working memory. *Science*, 319(5869), 1543–1546. https://doi.org/10.1126/science.1150769

Muhle-Karbe, P. S., Duncan, J., De Baene, W., Mitchell, D., & Brass, M. (2016). Neural coding for instruction-based task sets in human frontoparietal and visual cortex. *Cerebral Cortex*, 27, 1891–1905. http://doi.org/10.1093/cercor/bhw032

Muhle-Karbe, P. S., Myers, N. E., & Stokes, M. G. (2020). A hierarchy of functional states in working memory. *bioRxiv*, 51(2), 2020.04.16.044511. https://doi.org/10.1101/2020.04.16.044511

Myers, N. E., Chekroud, S. R., Stokes, M. G., & Nobre, A. C. (2018). Benefits of flexible prioritization in working memory can arise without costs. *Journal of Experimental Psychology: Human Perception and Performance*, 44(3), 398–411. https://doi.org/10.1037/xhp0000449

Myers, N. E., Rohenkohl, G., Wyart, V., Woolrich, M. W., Nobre, A. C., & Stokes, M. G. (2015). Testing sensory evidence against mnemonic templates. *eLife*, 4, 1–25. https://doi.org/10.7554/eLife.09000

Myers, N. E., Stokes, M. G., & Nobre, A. C. (2017). Prioritizing information during working memory: Beyond sustained internal attention. *Trends in Cognitive Sciences*, 21(6), 449–461. https://doi.org/10.1016/j.tics.2017.03.010

Nassar, M. R., Helmers, J. C., & Frank, M. J. (2018). Chunking as a rational strategy for lossy data compression in visual working memory. *Psychological Review*, 125(4), 486–511. https://doi.org/10.1037/rev0000101

Nee, D. E., & Jonides, J. (2014). Frontal-medial temporal interactions mediate transitions among representational states in short-term memory. *Journal of Neuroscience*, 34(23), 7964–7975. https://doi.org/10.1523/JNEUROSCI.0130-14.2014

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 411–421. https://doi.org/10.1037/0278-7393.28.3.411

Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review*, 124(1), 21–59. https://doi.org/10.1037/rev0000044

Olivers, C. N. L., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, 32(5), 1243–1265. https://doi.org/10.1037/0096-1523.32.5.1243

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*, 15(7), 327–334. https://doi.org/10.1016/j.tics.2011.05.004

Panichello, M. F., & Buschman, T. J. (2020). Selective control of working memory in prefrontal, parietal, and visual cortex. *bioRxiv*, 21, 2020.04.07.030718. https://doi.org/10.1101/2020.04.07.030718

Parthasarathy, A., Herikstad, R., Bong, J. H., Medina, F. S., Libedinsky, C., & Yen, S.-C. (2017). Mixed selectivity morphs population codes in prefrontal cortex. *Nature Neuroscience*, 1–16. https://doi.org/10.1038/s41593-017-0003-2

Remington, E. D., Narain, D., Hosseini, E. A., & Jazayeri, M. (2018). Flexible sensorimotor computations through rapid reconfiguration of cortical dynamics. *Neuron*, 98(5), 1005–1019.e5. https://doi.org/10.1016/j.neuron.2018.05.020

Rerko, L., & Oberauer, K. (2013). Focused, unfocused, and defocused information in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(4), 1075–1096. https://doi.org/10.1037/a0031172

Rerko, L., Souza, A. S., & Oberauer, K. (2014). Retro-cue benefits in working memory without sustained focal attention. *Memory & Cognition*, 42(5), 712–728. https://doi.org/10.3758/s13421-013-0392-8

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497(7451), 585–590. https://doi.org/10.1038/nature12160

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., & Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science*, 354(6316), 1136–1139. https://doi.org/10.1126/science.aah7011

Schneegans, S., & Bays, P. M. (2017). Restoration of fMRI decodability does not imply latent working memory states. *Journal of Cognitive Neuroscience*, 29(12), 1977–1994. https://doi.org/10.1162/jocn_a_01180

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 31(2), 248–261. https://doi.org/10.1037/0096-1523.31.2.248

Soto, D., Hodsoll, J., Rotshtein, P., & Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences*, 12(9), 342–348. https://doi.org/10.1016/j.tics.2008.05.007

Spaak, E., Watanabe, K., Funahashi, S., & Stokes, M. G. (2017). Stable and dynamic coding for working memory in primate prefrontal cortex. *Journal of Neuroscience*, 37(27), 6503–6516. https://doi.org/10.1523/JNEUROSCI.3364-16.2017

Spellman, T., Rigotti, M., Ahmari, S. E., Fusi, S., Gogos, J. A., & Gordon, J. A. (2015). Hippocampal-prefrontal input supports spatial encoding in working memory. *Nature*, 522(7556), 309–314. https://doi.org/10.1038/nature14445

Stokes, M. G. (2011). Top-down visual activity underlying VSTM and preparatory attention. *Neuropsychologia*, 49(6), 1425–1427. https://doi.org/10.1016/j.neuropsychologia.2011.02.004

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

Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, 78(2), 364–375. https://doi.org/10.1016/j.neuron.2013.01.039

Stroud, J. P., Porter, M. A., Hennequin, G., & Vogels, T. P. (2018). Motor primitives in space and time via targeted gain modulation in cortical networks. *Nature Neuroscience*, 1–15. https://doi.org/10.1038/s41593-018-0276-0

Sugase-Miyamoto, Y., Liu, Z., Wiener, M. C., Optican, L. M., Richmond, B. J., & Friston, K. J. (2008). Short-term memory trace in rapidly adapting synapses of inferior temporal cortex. *PLoS Computational Biology*, 4(5), e1000073. https://doi.org/10.1371/journal.pcbi.1000073

Tang, C., Herikstad, R., Parthasarathy, A., Libedinsky, C., & Yen, S.-C. (2019). Independent activity subspaces for working memory and motor preparation in the lateral prefrontal cortex. *bioRxiv*, 22, 756072. https://doi.org/10.1101/756072

Trübutschek, D., Marti, S., Ueberschär, H., & Dehaene, S. (2019). Probing the limits of activity-silent non-conscious working memory. *Proceedings of the National Academy of Sciences*, 116(28), 14358–14367. https://doi.org/10.1073/pnas.1820730116

Van der Stigchel, S., & Olivers, C. N. L. (2019). The flexible nature of the interaction between attention and working memory. *Journal of Cognition*, 2(1), 31. https://doi.org/10.5334/joc.68

van Loon, A. M., Olmos-Solis, K., Fahrenfort, J. J., & Olivers, C. N. (2018). Current and future goals are represented in opposite patterns in object-selective cortex. *eLife*, 7, 1427. https://doi.org/10.7554/eLife.38677

van Loon, A. M., Olmos-Solis, K., & Olivers, C. N. L. (2017). Subtle eye movement metrics reveal task-relevant representations prior to visual search. *Journal of Vision*, *17*(6), 13. https://doi.org/10.1167/17.6.13

van Moorselaar, D., Theeuwes, J., & Olivers, C. N. (2014). In competition for the attentional template: Can multiple items within visual working memory guide attention? *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1450–1464. https://doi.org/10.1037/a0036229

van Moorselaar, D., Theeuwes, J., & Olivers, C. N. L. (2016). Learning changes the attentional status of prospective memories. *Psychonomic Bulletin & Review*, *23*(5), 1483–1490. https://doi.org/10.3758/s13423-016-1008-7

Wang, X.-J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends in Neurosciences*, *24*(8), 455–463. https://doi.org/10.1016/S0166-2236(00)01868-3

Wimmer, K., Nykamp, D. Q., Constantinidis, C., & Compte, A. (2014). Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature Neuroscience*, *17*(3), 431–439. https://doi.org/10.1038/nn.3645

Wolff, M. J., Ding, J., Myers, N. E., & Stokes, M. G. (2015). Revealing hidden states in visual working memory using electroencephalography. *Frontiers in Systems Neuroscience*, *9*(5), 123. https://doi.org/10.3389/fnsys.2015.00123

Wolff, M. J., Jochim, J., Akyürek, E. G., & Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. *Nature Neuroscience*, *20*(6), 864–871. https://doi.org/10.1038/nn.4546

Yoo, S. B. M., & Hayden, B. Y. (2020). The transition from evaluation to selection involves neural subspace reorganization in core reward regions. *Neuron*, *105*(4), 712–724.e4. https://doi.org/10.1016/j.neuron.2019.11.013

Yu, Q., Teng, C., & Postle, B. R. (2020). Different states of priority recruit different neural representations in visual working memory. *PLoS Biology*, *18*(6), e3000769. https://doi.org/10.1371/journal.pbio.3000769

Zaksas, D., & Pasternak, T. (2006). Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. *Journal of Neuroscience*, *26*(45), 11726–11742. https://doi.org/10.1523/JNEUROSCI.3420-06.2006

Zhou, X., Katsuki, F., Qi, X.-L., & Constantinidis, C. (2012). Neurons with inverted tuning during the delay periods of working memory tasks in the dorsal prefrontal and posterior parietal cortex. *Journal of Neurophysiology*, *108*(1), 31–38. https://doi.org/10.1152/jn.01151.2011

Zucker, R. S., & Regehr, W. G. (2002). Short-term synaptic plasticity. *Annual Review of Physiology*, *64*(1), 355–405. doi:10.1146/annurev.physiol.64.092501.114547