Persistent Activity During Working Memory From Front to Back

Clayton E. Curtis[1,2]* and Thomas C. Sprague[3]

[1] *Department of Psychology, New York University, New York, NY, United States,* [2] *Center for Neural Science, New York University, New York, NY, United States,* [3] *Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA, United States*

Working memory (WM) extends the duration over which information is available for processing. Given its importance in supporting a wide-array of high level cognitive abilities, uncovering the neural mechanisms that underlie WM has been a primary goal of neuroscience research over the past century. Here, we critically review what we consider the two major "arcs" of inquiry, with a specific focus on findings that were theoretically transformative. For the first arc, we briefly review classic studies that led to the canonical WM theory that cast the prefrontal cortex (PFC) as a central player utilizing persistent activity of neurons as a mechanism for memory storage. We then consider recent challenges to the theory regarding the role of persistent neural activity. The second arc, which evolved over the last decade, stemmed from sophisticated computational neuroimaging approaches enabling researchers to decode the contents of WM from the patterns of neural activity in many parts of the brain including early visual cortex. We summarize key findings from these studies, their implications for WM theory, and finally the challenges these findings pose. Our goal in doing so is to identify barriers to developing a comprehensive theory of WM that will require a unification of these two "arcs" of research.

INTRODUCTION

The ability to store information for brief periods of time, so-called working memory (WM), is a building block for most of our higher cognitive functions, and its dysfunction is at the heart of a variety of psychiatric and neurologic symptoms. In the history of study into the neural mechanisms that support WM, an imperative goal of neuroscience, we would argue that there have been two main arcs. One began almost 50 years ago when Joaquin Fuster first reported that spiking measured from neurons in the macaque prefrontal cortex (PFC) persisted during a WM delay (Fuster and Alexander, 1971). Following this seminal publication, many researchers have measured this *persistent activity* with the goal of understanding how WM representations are stored by neural activity (Curtis and D'Esposito, 2003). The vast majority of the work has been focused on the PFC. The other arc began more recently, over the last decade, but has already made a tremendous impact on WM theory. Utilizing sophisticated computational neuroimaging approaches (e.g., machine learning, encoding models, etc.), researchers demonstrated that one can decode the contents of WM from the patterns of neural activity in early visual cortex (e.g., Harrison and Tong, 2009; Serences et al., 2009). This was surprising because at the time no existing data, and surely no WM theory, suggested that sensory cortices played a role in WM storage. The so-called *sensory recruitment*

theory of WM emerged from the ever-growing body of research suggesting a potential role for early visual cortex in visual WM. To a large extent, these two arcs have existed independently of one another. Here, we concisely review each arc along with challenges to the relevant theories, with the goal of identifying barriers that future research needs to address if an integrative theory of WM might develop.

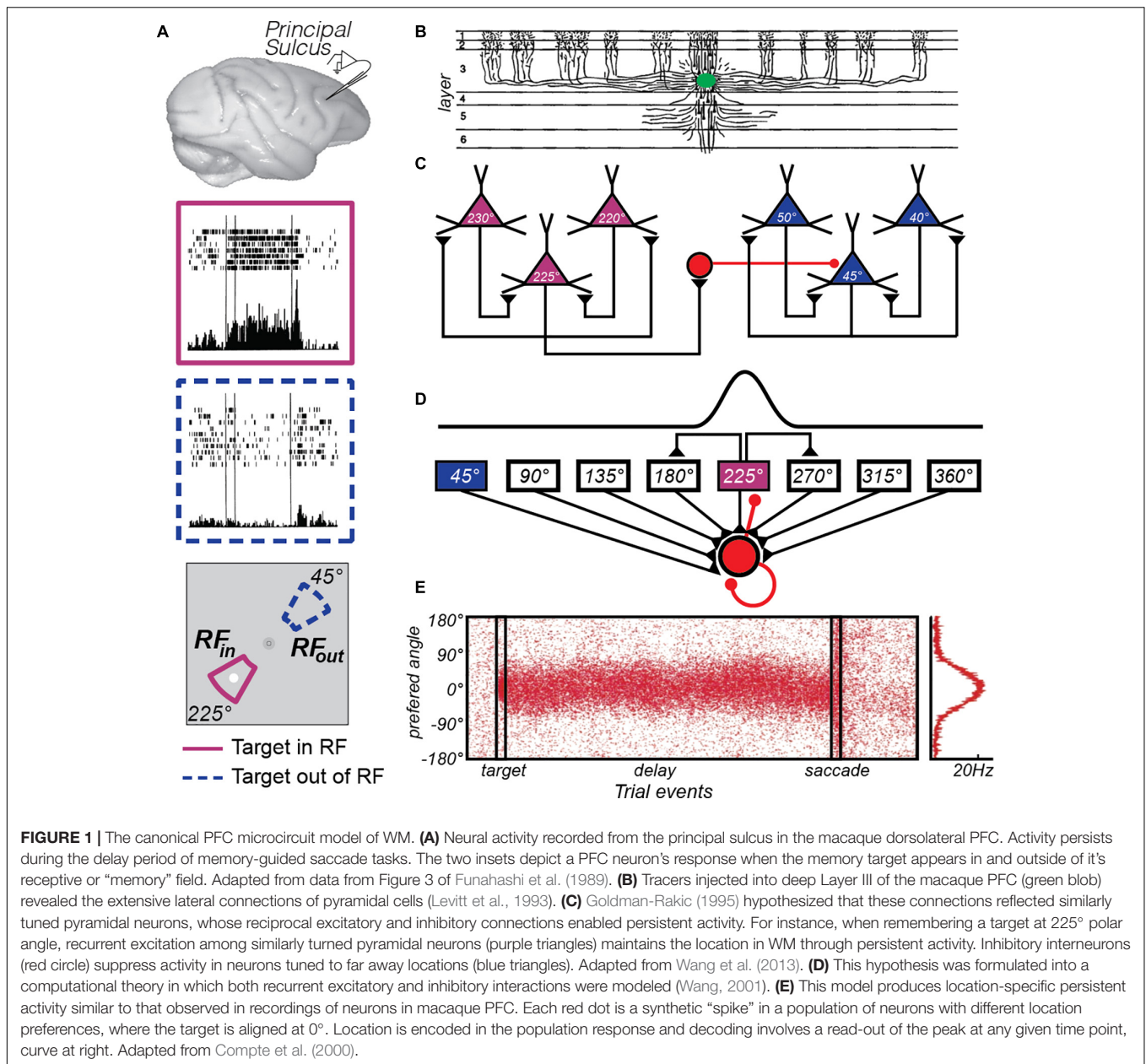## NEURAL ACTIVITY PERSISTS IN THE PREFRONTAL CORTEX

Following a century of studies investigating the effects of experimental lesions of the non-human primate cortex, researchers honed in on the principal sulcus in lateral PFC (from here on we will simply refer to this region as PFC) as a critical structure supporting WM functions (for a review see Curtis and D'Esposito, 2004). By 1971, an American lab (Fuster and Alexander, 1971) and a Japanese lab (Kubota and Niki, 1971) began recording extracellular neurophysiological signals from the PFC while macaques performed WM experiments. They reported that some neurons in the PFC tended to maintain an elevated rate of spiking, relative to pre-trial baseline firing rates, during WM retention intervals. Adapting an oculomotor version of the delayed response task, along with other experimental refinements, allowed Funahashi, Bruce, and Goldman-Rakic (Funahashi et al., 1989) to clarify several features of the persistent activity. First, they demonstrated that persistent activity in PFC neurons was memory stimulus selective in that, for a given neuron, it was typically restricted to one or two of the target positions in the contralateral hemifield (**Figure 1A**). This meshed well with a later report that experimental lesions of the PFC tended to impact memory for targets in the contralesional hemifield (Funahashi et al., 1993a). Second, they demonstrated that activity persisted for the duration of the memory delay (3 or 6 s) consistent with a mechanism that bridged the time between the past sensory event and the contingent behavior. Third, they demonstrated that the amplitude of persistent activity was reduced prior to memory errors. Because these features align with our notions of memory so closely, persistent activity was embraced as the neural basis of WM. It is no wonder, then, that the discovery of persistent activity is considered the most important scientific observation with regard to the neural mechanisms of WM. This now classic finding has been replicated numerous times and has had a tremendous impact on WM theory and how we study WM experimentally (as reviewed in Riley and Constantinidis, 2015).

Following these pioneering studies, the experimental techniques matured and over the next 30 years our knowledge about the relationships between persistent activity and WM accumulated. For example, persistent activity in PFC neurons is not limited to spatial WM. PFC neurons that show preferences for both simple (e.g., color) and complex (e.g., face) objects exhibit activity that persists while monkeys maintain these objects in WM (Quintana et al., 1988; Miller et al., 1996; Ó Scalaidhe et al., 1999; Fuster et al., 2000; Panichello and Buschman, 2021). Assuming that stimulus selective persistent

activity is the mechanism by which WM representations are stored, PFC neurons appear to store any type of stimulus feature, including the frequency of tactile flutter (Romo et al., 1999), the direction of dot motion (Zaksas and Pasternak, 2006; Mendoza-Halliday et al., 2014), sound location (Fuster et al., 2000; Kikuchi-Yorioka and Sawaguchi, 2000), and audiovisual macaque vocalizations (Hwang and Romanski, 2015). Moreover, they encode memory-guided prospective motor plans (Funahashi et al., 1993b; Takeda and Funahashi, 2002; Markowitz et al., 2015) and the prospective sensory features of a delayed paired associate (Rainer et al., 1999; Fuster et al., 2000). Finally, persistent activity appears to even encode complex task rules and contexts (Asaad et al., 2000; Wallis et al., 2001), abstract categories (Freedman et al., 2001), and selective conjunctions of objects and locations (Rao et al., 1997; Rainer et al., 1998) that cannot be explained by simpler stimulus or location specific representations.

CANONICAL PFC MICROCIRCUIT MODEL OF WM

Once the link between persistent activity in the PFC and WM was firmly established, many focused on determining the properties of neurons and circuits in the PFC that give rise to memory selective persistent activity. Pyramidal neurons in layer III of the PFC make horizontal connections with clusters of other pyramidal neurons in regular intervals (Levitt et al., 1993; Lund et al., 1993; Kritzer and Goldman-Rakic, 1995; **Figure 1B**). V1 neurons have a similar patchy horizontal connectivity (Gilbert and Wiesel, 1983) and connected neurons are more likely to have similar orientation tuning (Gilbert and Wiesel, 1989). By logic of induction, from these observations Goldman-Rakic theorized that similarly tuned (i.e., for location) pyramidal neurons in layer III are the source of glutamatergic excitatory recurrent connections that give rise to persistent activity (Goldman-Rakic, 1995; **Figure 1C**). Indeed, the persistent activity of PFC neurons with similar visuospatial tuning are correlated (Constantinidis et al., 2001). These excitatory dynamics are thought to be balanced by closely synchronized fast spiking inhibitory interneurons (Constantinidis and Goldman-Rakic, 2002), whose lateral inhibition is theorized to additionally help sculpt the spatial tuning of PFC pyramidal neurons (Rao et al., 2000). Goldman-Rakic's theory was formalized into a computational model that specified how excitatory recurrent activity, balanced and tuned by inhibition, could give rise to memory-specific persistent activity within a PFC microcircuit (Compte et al., 2000; Wang, 2001; **Figures 1D,E**). This theoretical model highlighted the importance of the slow kinetics of NMDA receptors, compared to the faster kinetics of AMPA receptors (Wang, 1999). Empirical evidence has generally supported many aspects of the PFC microcircuit model of WM. Persistent activity depends on glutamatergic synapses on long, thin spines connecting PFC neurons in layer III (Wang et al., 2011), and these excitatory currents depend on the slow kinetics of NMDA receptors to support persistent activity (Wang et al., 2013). Moreover, the model hypothesizes that small random drifts in the bumps of activity cause the seemingly random inaccuracies

FIGURE 1 | The canonical PFC microcircuit model of WM. **(A)** Neural activity recorded from the principal sulcus in the macaque dorsolateral PFC. Activity persists during the delay period of memory-guided saccade tasks. The two insets depict a PFC neuron's response when the memory target appears in and outside of it's receptive or "memory" field. Adapted from data from Figure 3 of Funahashi et al. (1989). **(B)** Tracers injected into deep Layer III of the macaque PFC (green blob) revealed the extensive lateral connections of pyramidal cells (Levitt et al., 1993). **(C)** Goldman-Rakic (1995) hypothesized that these connections reflected similarly tuned pyramidal neurons, whose reciprocal excitatory and inhibitory connections enabled persistent activity. For instance, when remembering a target at 225° polar angle, recurrent excitation among similarly turned pyramidal neurons (purple triangles) maintains the location in WM through persistent activity. Inhibitory interneurons (red circle) suppress activity in neurons tuned to far away locations (blue triangles). Adapted from Wang et al. (2013). **(D)** This hypothesis was formulated into a computational theory in which both recurrent excitatory and inhibitory interactions were modeled (Wang, 2001). **(E)** This model produces location-specific persistent activity similar to that observed in recordings of neurons in macaque PFC. Each red dot is a synthetic "spike" in a population of neurons with different location preferences, where the target is aligned at 0°. Location is encoded in the population response and decoding involves a read-out of the peak at any given time point, curve at right. Adapted from Compte et al. (2000).

in memory (Compte et al., 2000); but see (Standage and Paré, 2018). Evidence for this hypothesis exists, as clockwise or counterclockwise biases in population estimates of delay activity in macaque PFC neurons predict small angular errors in memory (Wimmer et al., 2014).

There are also anatomical properties that suggest advantages that PFC may have in its capacity for WM storage. These slow NMDA receptors are densely expressed in PFC, especially when compared to V1 (Wang et al., 2008). Pyramidal neurons in PFC, again compared to visual cortex, have larger and more complex dendritic branching with a greater number of spines (Oga et al., 2017), have more extensive horizontal collaterals in Layers II and III (Kritzer and Goldman-Rakic, 1995), and are twice as likely to form reciprocal connections (Wang et al., 2006). Together, the

excitatory connections theorized to form positive feedback loops to sustain WM representations (Goldman-Rakic, 1995) may be better supported by these anatomical features in PFC.

TRANSLATING THE PRIMATE PFC MODEL OF HUMAN WM

The success and impact of any animal model of human cognition depends on how well it translates to the species it is meant to model. It is not surprising then, that when brain imaging methods became widely available, researchers immediately predicted that indirect measures of neural activity could be used to measure persistent activity during WM in the homologous part of the

human PFC. It turned out not to be so easy. The first human brain imaging study of spatial WM *failed* to find that blood flow measured with Positron Emission Tomography (PET) localized to the dorsolateral PFC (Jonides et al., 1993). Then, the failure of several studies to find spatial WM-related delay period activity in the presumed homologous part of human dorsolateral PFC became the norm rather than the exception (Smith et al., 1996; Courtney et al., 1998; Zarahn et al., 1999; Rowe et al., 2000). A subsequent functional magnetic resonance imaging (fMRI) study from Goldman-Rakic's own lab succeeded in evoking dorsolateral PFC activity but only after increasing the WM load to five items (Leung et al., 2002). At the time it was assumed that fMRI did not have enough sensitivity to reliably measure persistent activity associated with maintaining a single item in WM. However, as we describe next this is unlikely the case and suggests alternative explanations.

Measuring neural activity with fMRI while humans perform spatial WM tasks, including memory-guided saccade WM tasks like those used to initially study the macaque PFC (Funahashi et al., 1989), we find that fMRI is perfectly sensitive to WM representations of single items (Curtis et al., 2004; Curtis and D'Esposito, 2006; Schluppeck et al., 2006; Srimal and Curtis, 2008; Tark and Curtis, 2009; Jerde et al., 2012; Sprague et al., 2014; Saber et al., 2015; Rahmati et al., 2020; Hallenbeck et al., in press). However, in none of the above cited studies did we find evidence that neural activity persists in the human dorsolateral PFC during simple spatial WM tasks. On the other hand, in each one of those studies we found evidence that activity persists, in a variety of meaningful ways, in the superior spur of the precentral sulcus (PCS) in the frontal cortex and/or in the posterior part of the intraparietal sulcus (IPS) (**Figure 2A**).
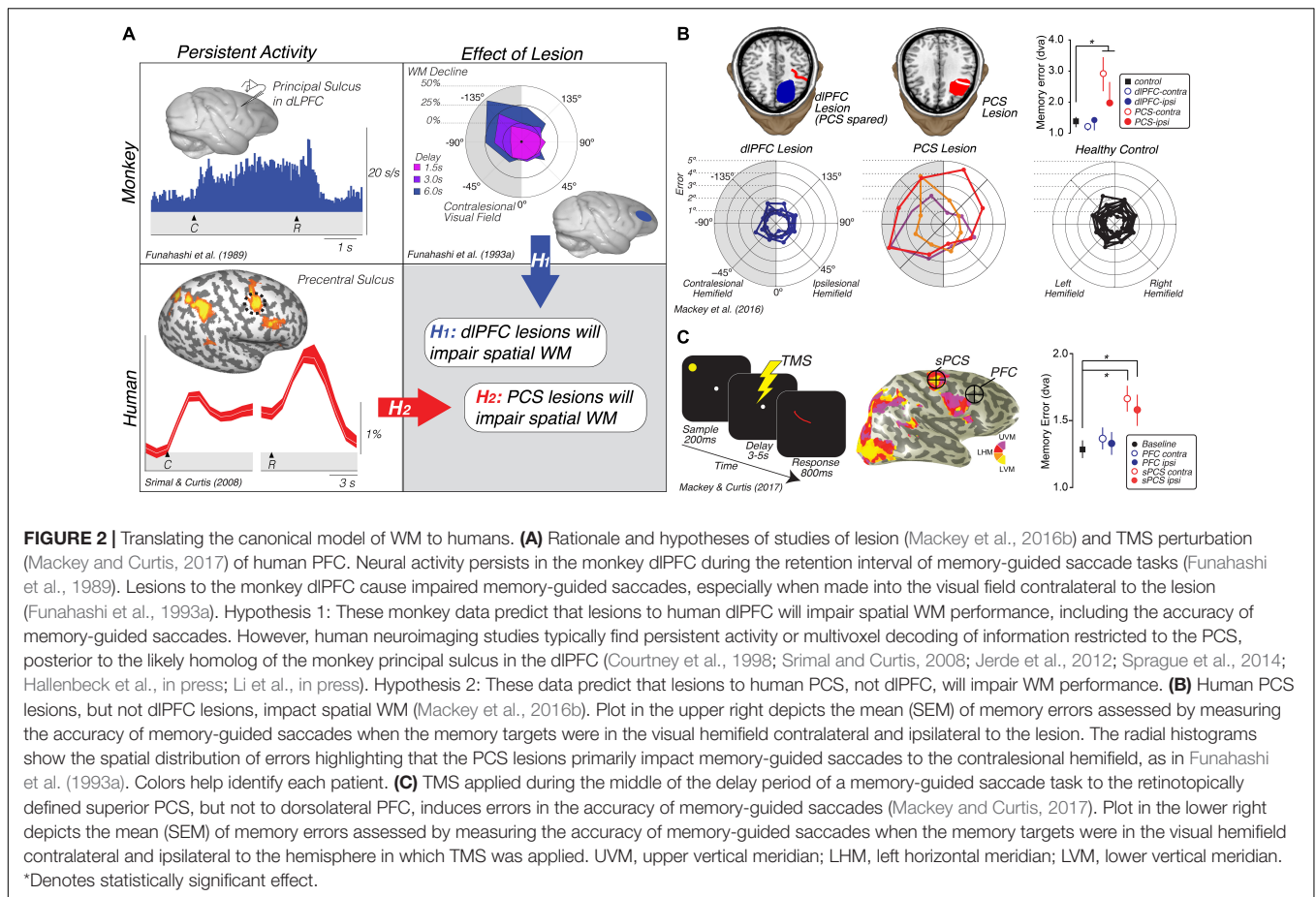
On the face of it, these results conflict between the two species. In the monkey, neurons in dorsolateral PFC show persistent activity and lesions cause WM impairments. However, in humans neural activity only persists in the PCS, not in more anterior parts of the PFC in areas homologous to the macaque principal sulcus. We generated two hypotheses to explain these conflicting results based on the impact that lesions to the PFC and PCS had on WM (**Figure 1A**). If lesions to the dorsolateral PFC that spare the PCS cause WM impairments, like they do in monkeys, this would indicate that fMRI may not be sensitive enough to measure persistent activity in that part of the brain (hypothesis 1). On the other hand, if lesions to the PCS, rather than PFC, cause WM impairments, this would indicate that the human dorsolateral PFC is not necessary for WM like it is in the monkey (hypothesis 2). In support of hypothesis 2, the accuracy of memory-guided saccades was unimpacted by dorsolateral PFC resections as long as they spared the PCS (**Figure 2B**). PCS lesions increased the magnitude of memory errors largely when the target was in the contralesional hemifield (Mackey et al., 2016b, 2017; Mackey and Curtis, 2017). In order to rule out other factors, like reorganization or compensation in the lesion patients, we repeated the study using transcranial magnetic stimulation (TMS) applied to the superior PCS and the intermediate frontal sulcus in the PFC during the memory delay in a healthy cohort of participants (Mackey et al., 2016b, 2017; Mackey and Curtis, 2017). The TMS results replicated the

patient study; TMS to the sPCS, but not dorsolateral PFC, caused an increase in memory-guided saccade errors (**Figure 2C**). These results are consistent with previous studies that have investigated the impact of dorsolateral PFC and/or PCS damage on both spatial and non-spatial forms of WM (D'Esposito and Postle, 1999; Ploner et al., 1999; Postle et al., 2003).

The question of what BOLD is actually measuring is an important question to consider with respect to persistent activity. In the highly influential paper by Logothetis et al. (2001), they reported that BOLD signals in V1 correlated highly with local field potentials (LFP) ($r^2 = 0.91$). However, what is less often recalled is that BOLD also correlated with spiking (multiunit activity (MUA); $r^2 = 0.73$). At longer timescales, when the MUA returned to pre-stimulus levels (i.e., adapted) in V1 despite the enduring visual stimulation, the LFP remained above baseline. Even in these conditions, the BOLD signal correlated with LFP ($r^2 = 0.52$), and almost as strongly with MUA ($r^2 = 0.45$). Despite this, many still believe that BOLD signals do not correlate with spiking, but instead only reflect the local processing of inputs to the region. What is correct depends on the question. If one is interested in the physiological coupling itself–specifically, the neural causes of the BOLD signal–then yes, the major driver of the BOLD response is afferent processing, which is better indexed by LFPs. Indeed, if one blocks the usually strong coupling between MUA and LFP with the use of a serotonin-agonist to hyperpolarize afferent membranes and thus block output spiking, the BOLD signal still correlates with LFP, but not MUA (Rauch et al., 2008). However, without such unnatural pharmacological interventions both MUA and LFP are both good predictors of the BOLD signal. Logothetis and Wandell (2004), made this point clearly in a review of the nature of the BOLD signal:

> "*In general, LFPs and MUA vary in a similar manner. Hence, at those sites where the LFPs predicted the BOLD response, the MUA did too. Across cortical sites there was a tendency for the LFP-based estimate to perform slightly better than the MUA-based estimate: The LFP signal predicted 7.6% more of the variance than the MUA. The difference, although small, was statistically significant. The larger variability of MUA was mostly attributable to the stronger adaptation effects observed in this frequency range of the mEFP [mean extracellular field potential].*" (pg. 747).

If the decoupling between BOLD and spiking is most affected by the adaptation of firing rate, then in the case of WM, one would predict a particularly strong coupling between BOLD and the persistent spiking of neurons (which do not show strong evidence for adaptation during delay periods, e.g., Funahashi et al., 1989). As a result, regions with persistent spiking activity should show strong BOLD signals. Thus, if persistent spiking activity in PFC supports WM in humans, in our view, this should be readily detectable in the BOLD signal measured with fMRI. This is true even if the correlation is indirect through changes in LFP power (Pesaran et al., 2002). Importantly, the observation of no persistent BOLD activation in PFC regions during tasks known to recruit persistent spiking in similar regions of macaques remains meaningful, and is strongly suggestive that these regions may not play a similar role in humans.

FIGURE 2 | Translating the canonical model of WM to humans. (A) Rationale and hypotheses of studies of lesion (Mackey et al., 2016b) and TMS perturbation (Mackey and Curtis, 2017) of human PFC. Neural activity persists in the monkey dlPFC during the retention interval of memory-guided saccade tasks (Funahashi et al., 1989). Lesions to the monkey dlPFC cause impaired memory-guided saccades, especially when made into the visual field contralateral to the lesion (Funahashi et al., 1993a). Hypothesis 1: These monkey data predict that lesions to human dlPFC will impair spatial WM performance, including the accuracy of memory-guided saccades. However, human neuroimaging studies typically find persistent activity or multivoxel decoding of information restricted to the PCS, posterior to the likely homolog of the monkey principal sulcus in the dlPFC (Courtney et al., 1998; Srimal and Curtis, 2008; Jerde et al., 2012; Sprague et al., 2014; Hallenbeck et al., in press; Li et al., in press). Hypothesis 2: These data predict that lesions to human PCS, not dlPFC, will impair WM performance. (B) Human PCS lesions, but not dlPFC lesions, impact spatial WM (Mackey et al., 2016b). Plot in the upper right depicts the mean (SEM) of memory errors assessed by measuring the accuracy of memory-guided saccades when the memory targets were in the visual hemifield contralateral and ipsilateral to the lesion. The radial histograms show the spatial distribution of errors highlighting that the PCS lesions primarily impact memory-guided saccades to the contralesional hemifield, as in Funahashi et al. (1993a). Colors help identify each patient. (C) TMS applied during the middle of the delay period of a memory-guided saccade task to the retinotopically defined superior PCS, but not to dorsolateral PFC, induces errors in the accuracy of memory-guided saccades (Mackey and Curtis, 2017). Plot in the lower right depicts the mean (SEM) of memory errors assessed by measuring the accuracy of memory-guided saccades when the memory targets were in the visual hemifield contralateral and ipsilateral to the hemisphere in which TMS was applied. UVM, upper vertical meridian; LHM, left horizontal meridian; LVM, lower vertical meridian. *Denotes statistically significant effect.

Now we return to trying to understand the discrepancy in WM findings between the human and monkey studies, and in doing so we need to consider a variety of possible explanations. First, a single item WM task may be too easy for humans relative to monkeys. Similar to the load argument discussed above (Leung et al., 2002), perhaps increasing the number of items increases the difficulty and thus recruits the human PFC. Nonetheless, the canonical WM theory does not specify that the dorsolateral PFC is only needed when the WM system is taxed with a challenging task. Moreover, other control processes such as reorganization and compression are necessary when one must maintain a number of items in WM (Rypma et al., 2002), especially when these approach or surpass capacity limits (Cowan, 2001). Second, perhaps a poor understanding of the homologies in either brain structure or function between the two species is more complicated than thought (Petrides et al., 2012). Third, the percentage of neurons in the macaque principal sulcus that show delay period activity is low (~10%) relative to the percentage of neurons in the frontal eye field (FEF) (~%50), located down in the anterior bank of the arcuate sulcus, and the lateral intraparietal (LIP) area (~%50). Due to very large receptive fields (RFs), the tuning for location during the memory delay is coarse in the PFC relative to the FEF and LIP (Mohler et al., 1973; Blatt et al., 1990; Hamed et al., 2001). Plus, the horizontally connected clusters of pyramidal neurons in layer III of the PFC form stripes that are spaced 0.2–0.8 mm apart (Kritzer and Goldman-Rakic, 1995). Perhaps fMRI is insensitive because this spatial separation dilutes over voxels the signal from an already small percentage of poorly tuned neurons persisting in the PFC. Fourth, there are surely true differences between the two species that cannot be attributed to the methods with which neural activity is measured. If we were considering rodent models of WM (e.g., Goard et al., 2016; Inagaki et al., 2019), we would be less bothered by possible mismatches in the exact brain areas, and would instead focus on the advantages of the animal model to learn about the precise neural mechanisms. One potential implication is that the mechanisms described in the microcircuit model of WM might be more applicable to cortical areas other than the human PFC. Indeed, lesions to the macaque FEF and LIP, as well as homologous areas in the human brain both impair WM performance (Dias and Segraves, 1999; Gaymard et al., 1999; Li et al., 1999; Ploner et al., 1999; Mackey et al., 2016a,b; Mackey and Curtis, 2017).
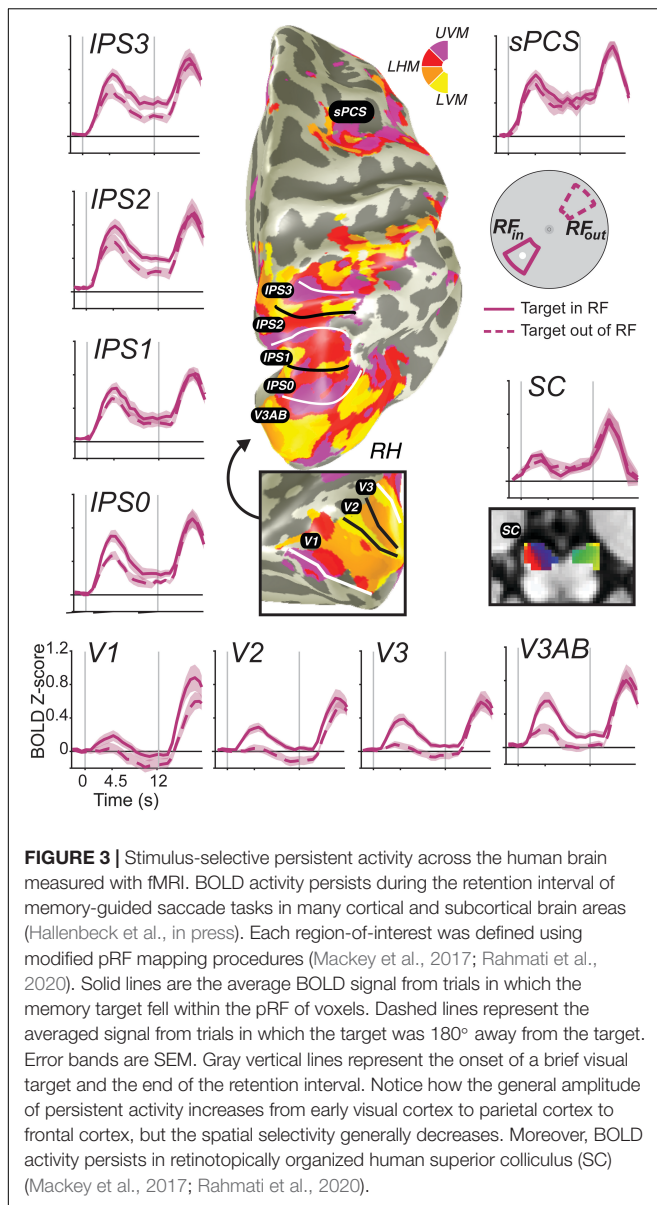
NEURAL ACTIVITY PERSISTS BEYOND PFC

The dorsolateral PFC is not the only brain area housing neurons that persist during WM (Leavitt et al., 2017).

Funahashi et al. (1989) also reported that neurons in the FEF showed spatially tuned persistent activity. As mentioned above, persistent activity is more common among FEF neurons than PFC, more robust, and more spatially selective (Goldberg and Bruce, 1985; Sommer and Wurtz, 2001; Merrikhi et al., 2017; Hart and Huk, 2020). Activity persists during WM tasks in several other frontal areas including the dorsal premotor cortex (PMD) (Rossi-Pool et al., 2017; Bastos et al., 2018), the supplementary eye fields (SEFs) (Shichinohe et al., 2009; Fukushima et al., 2011), the anterior cingulate cortex (ACC) (Kamiński et al., 2017), and even the orbitofrontal cortex (OFC) (Ichihara-Takeda and Funahashi, 2007). Moreover, neurons in LIP and 7a also show spatially selective and robust persistent activity (Gnadt and Andersen, 1988; Barash et al., 1991; Constantinidis and Steinmetz, 1996; Chafee and Goldman-Rakic, 1998; Pesaran et al., 2002; Hart and Huk, 2020). In the temporal lobe, WM selective persistent activity has been reported in neurons in monkey inferotemporal (IT) cortex (Fuster and Jervey, 1981; Miyashita and Chang, 1988; Miller et al., 1993; Chelazzi et al., 1998) and even in hippocampus and nearby entorhinal/perirhinal cortex (Miller and Desimone, 1994; Suzuki et al., 1997; Wirth et al., 2003). Evidence also exists that persistent neuronal activity carries sensory information about object identity in V4 (Hayden and Gallant, 2013) and motion in area MT (Bisley et al., 2004). However, these results are controversial (Pasternak and Greenlee, 2005; Leavitt et al., 2017) as other studies have reported an absence of persistent activity among neurons in MT coding for the remembered motion direction (Mendoza-Halliday et al., 2014) and the persistent activity may be limited to the early and late phases of the delay (Bisley et al., 2004). Whether activity in individual neurons persists in these sensory areas may depend on the type of representational format an animal might be using to store the memory as opposed to the representation formed during perception. For instance, it is unlikely that memory for dot motion is a replay of hundreds of dots moving over time. Perhaps, that temporally evolving percept is compressed or recoded into something like a single directional vector that does not drive MT. Remarkably and surprising to many neuroscientists, even neurons in V1 show activity which persists during WM delays (Supèr et al., 2001; van Kerkoerle et al., 2017). Finally, neurons in subcortical areas like the superior colliculus (SC) (Shen et al., 2011; Dash et al., 2015; Sadeh et al., 2018) and mediodorsal thalamus (Funahashi, 2013) are spatially tuned and carry location information during WM delay periods. The point we are trying to make in this section is that the presumed mechanism that supports WM–memoranda-specific persistent activity–is not exclusively localized to the dorsolateral PFC. Rather it appears to be a mechanism used by many parts of the brain to encode enduring representations useful for memory-guided decisions.

In humans, measuring delay period activity with fMRI BOLD supports these non-human primate reports in just how widely distributed persistent activity appears to be during WM. There have been a number of reviews recently of human neuroimaging studies of WM (e.g., Sreenivasan et al., 2014; D'Esposito and Postle, 2015; Christophel et al., 2017; Sreenivasan and D'Esposito, 2019) and thus we will instead focus on instructive examples of persistent activity measured in humans

with fMRI. Moreover, we focus on spatial WM because of its widespread use in both species and the ease with which neural encoding properties can be measured in both species. In many of the monkey electrophysiological studies reviewed above, an important first step involved characterizing each neuron's RF or its preferred stimulus feature. This then allowed researchers to compare memory responses between stimuli placed within and outside of each neuron's RF (or compare between preferred and non-preferred stimuli). Utilizing the same logic, advances in population receptive field (pRF) mapping (Dumoulin and Wandell, 2008; Wandell and Winawer, 2015; Mackey et al., 2017) allow researchers to compare BOLD estimates of persistent activity between trials in which the memoranda fall within and outside of a voxel's pRF. In **Figure 3**, the time courses of BOLD activity during a memory-guided saccade task are shown for ten visual field maps (Rahmati et al., 2020; Hallenbeck et al., in press). Each visual field map contains either an upright or inverted representation of the contralateral visual field. Within each map, the location and size of the pRF of each voxel in these maps can be estimated using non-linear optimization techniques. Then, averaging BOLD signal over trials can be performed in a principled way according to the match between voxels' pRF positions and the locations of the memorized targets. Overall, activity persists during the delay period in almost all of these maps. Moreover, the amplitude of BOLD activity is generally greater among voxels with pRFs matching the target, compared to voxels with pRFs 180° away from the target (on the opposite side of fixation).

Based on these data, two major gradients can be seen. One, the overall amplitude of persistent activity increases moving up the visual hierarchy from visual cortex to parietal cortex to frontal cortex. Two, the spatial selectivity (difference between when the target is in or out of the voxel's pRF) generally decreases up that same hierarchy. Even within the parietal cortex, we can see both of these gradients from IPS0 to IPS3. Of special interest, many neuroimaging studies have failed to find persistent activity in V1 [or in early visual cortex for that matter; e.g., (Ester et al., 2009; Harrison and Tong, 2009; Offen et al., 2009; Serences et al., 2009; Riggall and Postle, 2012; Albers et al., 2013)]. These studies, however, typically averaged over all voxels in V1, likely missing the more localized activity that persists associated with a given remembered stimulus. Note how the voxels in V1 with pRFs overlapping the small memory target showed a brief transient response time-locked to the target stimulus, but activation does not remain above the pre-trial baseline for the entire delay period. On the one hand, V1 does not meet the strict definition of persistent activity. On the other hand, if one considers the relationship between encoding and decoding from neural populations, then it does meet the definition of persistent activity. Namely, the WM representation is clearly encoded in the population as evident by the difference between the two time-courses. Moreover, the same decoder applied to read-out the population response would recover the target location despite the average signal dipping back down to pre-trial levels. We reported the same pattern in V1 previously even on trials in which the location of the visual target was different from the location of the memory-guided saccade, using an antisaccade

FIGURE 3 | Stimulus-selective persistent activity across the human brain measured with fMRI. BOLD activity persists during the retention interval of memory-guided saccade tasks in many cortical and subcortical brain areas (Hallenbeck et al., in press). Each region-of-interest was defined using modified pRF mapping procedures (Mackey et al., 2017; Rahmati et al., 2020). Solid lines are the average BOLD signal from trials in which the memory target fell within the pRF of voxels. Dashed lines represent the averaged signal from trials in which the target was 180° away from the target. Error bands are SEM. Gray vertical lines represent the onset of a brief visual target and the end of the retention interval. Notice how the general amplitude of persistent activity increases from early visual cortex to parietal cortex to frontal cortex, but the spatial selectivity generally decreases. Moreover, BOLD activity persists in retinotopically organized human superior colliculus (SC) (Mackey et al., 2017; Rahmati et al., 2020).

procedure (Saber et al., 2015), indicating that the response is memory related and not solely a residual BOLD response due to the visual transient. These advances allow for promising and more direct comparisons between monkey electrophysiology and human neuroimaging.

DECODING WM CONTENTS FROM POPULATION-LEVEL ACTIVATION PATTERNS

As reviewed above in the section on translational studies, attempts to identify persistent activity in human dorsolateral PFC during WM tasks that require simple maintenance have largely been unsuccessful. However, these studies typically leveraged

mass univariate analysis approaches which average responses across all trials in the experiment and over neighboring voxels by way of smoothing. As a result, these analyses effectively focus on the *similarities* in fMRI activation across all unique remembered stimuli. That is–to isolate activation related to the maintenance of information over the delay period, trials corresponding to all possible WM contents are combined. Such averaging necessarily masks important differences in activation associated with specific types of stimuli–for example, the particular location, orientation, or color held in WM.

Beginning at the turn of the century, human neuroimaging researchers began considering the possibility that *patterns* of brain activation measured with fMRI could discriminate between different stimulus or task conditions, rather than only considering elevated or suppressed *average* activation (Haxby et al., 2001; Norman et al., 2006). These methods primarily involve "decoding" which of several stimuli was present using machine learning tools, such as support vector machines. When these methods were turned to the early visual system, they demonstrated a remarkable ability to decode which orientation was viewed based on visual cortex activation patterns (Haynes and Rees, 2005; Kamitani and Tong, 2005). This was a surprising feat–the anatomical organization of orientation selectivity in the early visual system was thought to be too fine for study with the relatively coarse spatial resolution of fMRI (on the order of 2–3 mm per voxel). However, because the coarse sampling of the fine orientation columns is imperfect and uneven, the observed pattern of activation differed across stimulus orientations, enabling the decoding algorithm to detect these subtle differences and accurately decode which orientation was viewed (Boynton, 2005; Swisher et al., 2010). It should be noted that there exists considerable skepticism about the exact signals driving successful orientation decoding performance in these studies (Freeman et al., 2011, 2013; Alink et al., 2013; Carlson, 2014; Maloney, 2015; Pratte et al., 2016; Roth et al., 2018). Regardless of the source of the signals, it remains possible to recover distinctions in brain activation patterns associated with visual stimulus features.

Soon thereafter, these methods were applied to WM: Harrison and Tong (2009) and Serences et al. (2009) each reported success applying similar decoding techniques to visual cortex fMRI activation patterns measured during the delay-period of WM tasks. In each case, the authors demonstrated that only remembered information could be decoded, and non-remembered information [e.g., a discarded feature (Serences et al., 2009), or a discarded stimulus (Harrison and Tong, 2009)] was not maintained, demonstrating that these results cannot only be due to lingering sensory-evoked activation present in the slow hemodynamic signals measured with fMRI. This pair of studies offered convincing evidence for an important role of early sensory regions in supporting WM representations, especially when the features to be maintained are well-represented within those regions. This *sensory recruitment model* of WM posits that the previously identified sustained delay-period activation observed in association cortex acts to coordinate stimulus-specific representations in sensory cortex (Curtis and D'Esposito, 2003; Postle, 2006; D'Esposito and Postle, 2015; Serences, 2016).

In the decade since, dozens of studies have applied similar methods to decode visual stimulus features such as orientation, motion direction, color, spatial position, and the identity of a spatial pattern from brain activation patterns measured from striate and extrastriate visual cortex (Christophel et al., 2017). Moreover, modified versions of these decoding methods, including cvMANOVA (Allefeld and Haynes, 2014; Christophel et al., 2018a), inverted encoding models (IEMs) (**Figure 4A**; Ester et al., 2013; Sprague et al., 2014), and Bayesian decoding methods (van Bergen et al., 2015; van Bergen and Jehee, 2018, 2021; Brissenden et al., 2021; Li et al., in press) have increasingly improved the resolution and sensitivity of these methods to differences between conditions, and, ultimately, between individual trials. These new methods have revealed feature-selective representations broadly across visual, parietal, and frontal cortex (Christophel et al., 2012, 2018a,b; Jerde et al., 2012; Christophel and Haynes, 2014; Sprague et al., 2014; Ester et al., 2015; Yu and Shim, 2017; Rahmati et al., 2018; Li et al., in press), along with subcortical regions including the SC (Rahmati et al., 2020) and cerebellum (Brissenden et al., 2021). In human neuroimaging, evidence for stimulus-selective persistent activity abounds throughout the brain (**Figure 4B**). One challenge is to understand why there are so many WM representations distributed across the cortex. Perhaps they contain different formats of WM useful for various sensory, motor, and cognitive functions. Conversely, they might reflect some representation that is shared with WM. For instance, we found that decoders trained to predict spatial locations during the delay of a spatial WM cross-predicted the locations on the other tasks (e.g., covert attention and saccade planning) in sPCS and in IPS2 (Jerde et al., 2012). This suggested that the delay period patterns of activity may be interchangeable across spatial WM, attention, and saccade planning, and may reflect a common representation akin to attentional priority (Serences and Yantis, 2004; Fecteau and Munoz, 2006; Zelinsky and Bisley, 2015). Therefore, future research needs to investigate what types of information are stored in persistent activity.

Interestingly, in many cases, when sustained delay-period activation is compared directly against stimulus-selective activation patterns, complementary results are found (Postle, 2015). As an example, Riggall and Postle (2012) compared univariate delay-period activation and decoded information content for regions responsive to visual stimuli and those with elevated responses during the delay period of a WM task. In the stimulus-responsive regions (which were primarily in extrastriate visual cortex), a decoding algorithm was able to successfully recover the direction of motion remembered by the participants, but these regions did not show elevated delay-period activation. Conversely, in delay period-responsive regions (which were primarily in the IPS dorsal frontal cortex), the authors could not decode the remembered stimulus value, but did observe sustained delay-period activation spanning the sample and the probe stimulus. In a subsequent study in which WM load was additionally manipulated, sustained delay-period activation in frontal and parietal regions was shown to increase as WM load increased from 1 to 3 items, while a similar change in average activation was not observed in sensory
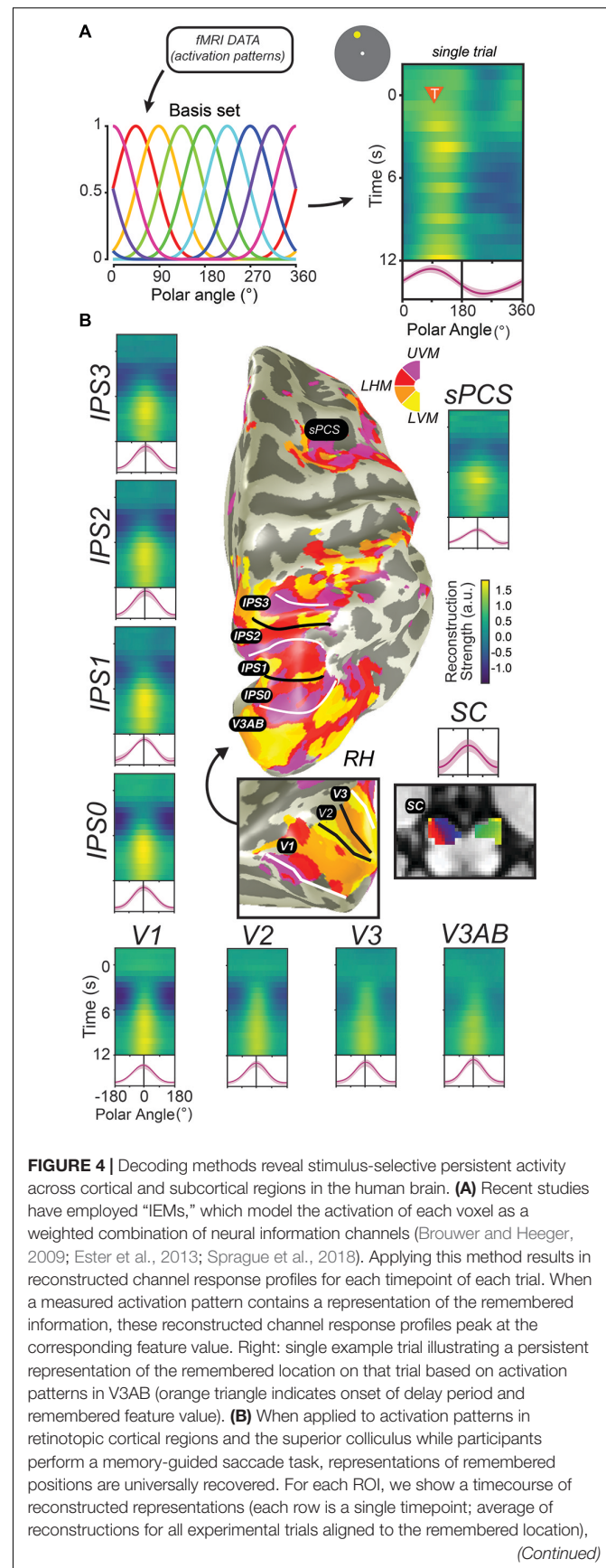

FIGURE 4 | Decoding methods reveal stimulus-selective persistent activity across cortical and subcortical regions in the human brain. **(A)** Recent studies have employed "IEMs," which model the activation of each voxel as a weighted combination of neural information channels (Brouwer and Heeger, 2009; Ester et al., 2013; Sprague et al., 2018). Applying this method results in reconstructed channel response profiles for each timepoint of each trial. When a measured activation pattern contains a representation of the remembered information, these reconstructed channel response profiles peak at the corresponding feature value. Right: single example trial illustrating a persistent representation of the remembered location on that trial based on activation patterns in V3AB (orange triangle indicates onset of delay period and remembered feature value). **(B)** When applied to activation patterns in retinotopic cortical regions and the superior colliculus while participants perform a memory-guided saccade task, representations of remembered positions are universally recovered. For each ROI, we show a timecourse of reconstructed representations (each row is a single timepoint; average of reconstructions for all experimental trials aligned to the remembered location),

(Continued)

regions (Emrich et al., 2013). However, the authors could reliably decode the remembered stimuli from activation patterns in sensory regions, with accuracy decreasing as WM load increased. Once again, this has been taken to suggest that regions showing elevated delay-period activity may not be those which represent the WM content itself, and that instead there may be a division of labor between frontal and parietal regions which help coordinate WM representations and sensory regions which encode stimulus values themselves (Postle, 2015; Postle and Yu, 2020).

However, when interpreting results from these decoding studies, it is critical to consider how these various algorithms operate to discriminate between remembered visual stimuli. The key feature of any decoding algorithm is that it identifies a reliable difference between activation patterns associated with different modeled stimulus values. While different approaches use different assumptions about the structure of these activation patterns and their noise covariance, this core feature remains. Accordingly, if a decoder can reliably pick up on differences between activation patterns within a region associated with different stimulus values, this necessarily means that some neurons (or, at least, signals resulting from neural activity) are more active than others in a reliable way. That is–the decoders aren't magic–they're just exploiting the structure of signals measured from neural tissue to optimally extract activation associated with different stimulus values. And, importantly, some stimulus values result in increased activation in some measured units, while other stimulus values result in increased activation in other measured units. As a trivial example, one could build a visual stimulus decoder based on a machine learning algorithm (e.g., support vector machines) to decode which side of the screen is stimulated by a large flickering checkerboard–a stimulus that is well-understood to evoke extremely strong and reliable fMRI signals in contralateral visual cortex. The decoder would perform extremely well–likely approaching 100% correct decoding performance. While in this case it wouldn't be necessary to apply the decoding algorithm to show that primary visual cortex encodes the retinotopic location of a stimulus, because a simple fMRI contrast would reveal strong evidence for such a result, this example remains illustrative: the decoder would be basing its judgment on localized increases in activation within a subset of the population of voxels.

When such an analysis is applied to data acquired during a memory-guided saccade task analogous to that used in macaques, greater activation is measured in voxels with spatial RFs near the remembered location as compared to those voxels with spatial RFs farther away (**Figure 3**; Saber et al., 2015; Hallenbeck et al., in press). Moreover, this holds for features like orientation: recent studies which have instead attempted to "localize" voxels preferring one or another orientation and directly compare activation between these subpopulations support this notion:

voxels labeled with the orientation remembered on a trial show elevated activation compared with those labeled with the non-remembered orientation (Lawrence et al., 2018). These results track with those observed in the classical studies of macaque DLPFC which show elevated neural firing for neurons which prefer the remembered location as compared to those with more distal preferences (Funahashi et al., 1989).

In our view, the ability to accurately decode which of several stimuli is held in WM is consistent with the definition of persistent activity: these results are driven (in large part) by different activity levels between different stimulus values during a WM delay period. Thus, decoding studies which observe stimulus-selective activation patterns in different cortical and subcortical brain regions should be considered to provide support for stimulus-selective persistent activity. Decoding of WM content and elevated delay-period activation may, in many cases, be considered two sides of the same coin (**Figure 3** vs. **Figure 4B**). Recent advances in decoding methods described above (IEM, cvMANOVA, and Bayesian generative models) have further extended the set of regions from which WM content can be decoded. Ester et al. (2015) and Yu and Shim (2017) applied IEMs to decode orientation and color from several parietal and prefrontal regions, and Christophel et al. (2018b) applied a non-parametric decoder based on a multivariate ANOVA to decode remembered orientation from the same regions from a large sample of fMRI participants ($n = 87$). Where sustained delay-period activation is found in humans, successful decoding of WM content seems to soon follow as the capabilities of methods advance.

FURTHER CHALLENGES TO THE CANONICAL PFC MODEL OF WM

So far we have described several findings that challenge the canonical WM model. In humans, simple WM does not depend on the dorsolateral PFC. Additionally, the persistent activity of neurons in PFC that sits at the heart of the canonical WM model is observed in many other brain regions, including early visual cortex. Together, these findings suggest that perhaps theories have tended to overemphasize the unique importance of PFC for WM. Additionally, further challenges have recently arisen to the very nature of what role persistent activity plays in WM.

Is Persistent Activity in PFC an Artifact of Averaging?

First, some have questioned whether persistent activity in PFC neurons is an artifact of averaging over trials (Shafi et al., 2007; Stokes and Spaak, 2016; Spaak et al., 2017). Similarly, the spiking activity of single PFC neurons might be best described as idiosyncratic bursts rather than persistent, and perhaps PFC activity is better characterized as "bubbles" of oscillations in LFP (Lundqvist et al., 2016, 2018). However, even if one accepts this to be the case, the original theoretical model does not need to be adjusted. The canonical model put forth by Goldman-Rakic (1995) and its later formalization as a computational model (Compte et al., 2000) never specified that WM representations

were stored by the persistent activity of single neurons. On the contrary, even the earliest versions of the model were inspired by the anatomy of layer III PFC neurons that were proposed to be clustered in pools of similarly tuned neurons with recurrent excitatory connections. Furthermore, the computational model clearly encodes WM representations through the overall activity of a population of neurons where the bump of activity could result from numerous and changing configurations of neurons. Perhaps the fact that the evidence for the theory took the form of averaged recordings of single neurons may have confused the issue. Nonetheless, the dynamics of single neurons involved in the population code deserves further investigation both at the empirical and theoretical level.

Dynamic Codes for WM Content

Second, some have questioned the temporal stability of WM representations encoded by the delay period activity of PFC neurons (Parthasarathy et al., 2017, 2019; Spaak et al., 2017; Cavanagh et al., 2018; Wasmuht et al., 2018). Based on analyses comparing the activity of groups of neurons across timepoints within trials, these studies have concluded that in some circumstances the population activity of PFC neurons that code for WM representations dynamically changes over time. This could be a real challenge to the canonical model of WM because this mechanism is at odds with a stable fixed activity pattern linking neuronal turning preferences with features stored in WM. Specifically, if WM representations were primarily dynamic, a downstream area would have to know about and track the dynamics of each neuron's encoding properties (its mnemonic tuning function as it unfolds over the trial) in order to read out the represented feature value from the population response at a given timepoint.

However, recent theoretical and empirical demonstrations have mitigated these concerns. Even when activity patterns are somewhat dynamic, such that the correlation between activity patterns is lower for points further separated in time than for points nearer in time, the population can be shown to have the same information content. Specifically, Murray et al. (2017) demonstrated that dynamic activity patterns that are occasionally observed in PFC exist within a "stable subspace" of the full population activity space, such that a downstream region could apply a fixed linear readout to accurately recover WM information throughout the delay period. Thus, at least in some cases, dynamic codes may only appear this way on the surface (Murray et al., 2017; Parthasarathy et al., 2019).

While there certainly does exist ample evidence that dynamic responses at the single-unit level can be observed, and that they can in some cases support a stable population-level neural code, it is critical to note that these studies do not negate the existence nor importance of other stable coding mechanisms, some of which are observed in the same studies. For example, it has been shown that neurons with dynamic responses and those with stable responses coexist in PFC, and their response dynamics can be well-predicted by their intrinsic "time constant" [their autocorrelation function measured from inter-trial intervals; (Wasmuht et al., 2018)]. That is–nearby neurons in the same brain region can either show evidence for dynamic coding

or stable coding. In another study, macaques performed an oculomotor delayed response task with an intervening irrelevant distractor stimulus. Activity patterns measured from LPFC "morphed" following the distractor, but patterns measured from the FEF of the same animals did not show evidence for such dynamic morphing (Parthasarathy et al., 2017). These results show that even when dynamic codes are observed, stable subspaces (consistent with a fixed readout rule) can account for a large amount of the response dynamics, and moreover, that stable coding is simultaneously observed in other neurons and/or brain regions.

Mixed Selectivity in PFC

Third, PFC neurons appear to have mixed selectivity as they can change their responsiveness to the same stimulus or behavioral response depending on subtle contextual changes within a task (Sigala et al., 2008; Machens et al., 2010; Mante et al., 2013; Rigotti et al., 2013). This could have several implications, including that the population response does not encode straightforward task variables, or that it encodes some latent variables that have yet to be discovered, or that it is dynamic over time at the timescale of the recording session. Nonetheless, there are some advantages to mixed selectivity. For example, the idea that the PFC can store any type of feature in WM implies that the entire manifold of encoding mechanisms housed in our sensory cortices might need to be duplicated just for short term storage, which seems highly inefficient at best. Mixed selectivity could vastly increase the encoding capacity of a given population (Rigotti et al., 2013). However, incorporating this concept into the canonical PFC model of WM would require altering the theory in ways that approach the way in which the hippocampus is thought to use mixed selectivity and sparse coding for long-term memory (Rolls and Treves, 1990; McClelland et al., 1995). Moreover, perhaps we have yet to discover the mechanisms by which the population response in PFC is demixed when it is readout by other brain areas (Machens, 2010).

"Activity-Silent" WM Representations

Fourth, metabolically economical models propose that persistent spiking may induce fast-timescale synaptic changes that encode stimulus properties that can be later retrieved efficiently via stimulus-agnostic "pinging" of the network (Mongillo et al., 2008; Stokes et al., 2013; Rose et al., 2016; Wolff et al., 2017), instructive cues (Lewis-Peacock et al., 2012; Sprague et al., 2016; LaRocque et al., 2017; Lorenc et al., 2020), and/or spontaneous internal neural reactivation signals (Lundqvist et al., 2016, 2018). In the empirical reports, decoding performance reliably drops around chance levels at one point in the trial, but a subsequent visual stimulus (Wolff et al., 2015, 2017, 2020a,b), TMS pulse (Rose et al., 2016), or task instruction (Lewis-Peacock et al., 2012; Sprague et al., 2016; LaRocque et al., 2017) results in a "reactivation" of an otherwise "latent" WM representation. This negative evidence, in the form of poor or at-chance decoding performance prior to reactivation, has been used to suggest that currently irrelevant information in WM is not maintained in an active state accessible to the measured neural signals fed into the decoding algorithm.

However, these studies additionally cannot rule out a key role for persistent activity in supporting WM behavior. When larger sample sizes and more sensitive analysis techniques are applied, there appears to be some positive evidence for representations of irrelevant WM information (Christophel et al., 2018b; Iamshchinina et al., 2021). While these positive results do not invalidate the previous negative observations, it does suggest that it can be possible–with a sufficient sample size–to find evidence for WM representations that elude studies with smaller sample sizes. The studies which have "pinged" human participants with irrelevant visual stimuli or TMS pulses can also not conclusively demonstrate that there existed *no information* prior to the reactivation stimulus. While the information may not have been accessible with EEG or MEG measurements, it may have existed as spontaneous oscillations in the electrophysiological recordings (LaRocque et al., 2013; Foster et al., 2016). A recent reanalysis of the data shown in Wolff et al. (2017) suggests this latter possibility (Barbosa et al., 2021). Finally, modeling has shown that observations of increased information content in IEM-based stimulus reconstructions following a task cue (Sprague et al., 2016) are not diagnostic of a transition from a passive to an active code (Schneegans and Bays, 2017). While this study (Sprague et al., 2016) found an enhancement in the decodable information about remembered spatial position following an informative cue, it remains the case that weak information may have been present prior to the cue, but was inaccessible to the fMRI signal and/or decoding algorithm employed (e.g., Christophel et al., 2018b).

Distractors Impact WM Representations in Sensory Regions

Cognitive theories about the nature of WM representations have long been informed by behavioral studies of the distracting effects of material presented during WM retention intervals (Baddeley, 1986). Intervening information is more disruptive when its features match the contents of WM. For instance, intervening phonological but not visual information impairs one's ability to maintain visually presented strings of letters, suggesting an important role of articulatory processes for items that are verbalizable (Logie et al., 1990). Similarly, intervening visuospatial processing and oculomotion selectively impacts spatial WM (Postle et al., 2006). In general, many conclusions about the formats of WM representations depend on the logic that the effectiveness of distraction depends on how well the representational formats of the distractor and memoranda are matched. Neuroscientific studies have also relied on a similar logic, assuming that the competition or interaction between the neural representations of the memoranda and the intervening distractor disrupts memory. As reviewed above, early electrophysiology and neuroimaging studies focused primarily on the importance of persistent activity in the PFC. Until recently, the potential importance of posterior cortical areas in WM had been largely neglected. Indeed, neural activity in monkey inferotemporal cortex is less robust during memory delays and the selectivity of activity appears to be disrupted by intervening distractors, while PFC representations appear resistant to distraction (Miller et al., 1996). Similarly,

memoranda-specific delay period activity of neurons in the monkey PFC resists the effect of distractors, especially when compared to neurons in posterior parietal cortex (di Pellegrino and Wise, 1993; Constantinidis and Steinmetz, 1996; Suzuki and Gottlieb, 2013). Inferences stemming from the underlying logic of these distractor studies imply that the PFC, rather than posterior cortical areas, is critical for WM storage.

Recent human neuroimaging studies have further addressed this issue using various decoding methods, with a primary focus on whether information about remembered features can be found in visual cortex in the presence of an intervening distracting stimulus. Bettencourt and Xu (2016) decoded orientations held in visual WM using activation patterns in visual and parietal cortex on trials with and without distracting visual stimuli during the delay period (faces and gazebos). When it was predictable whether a distractor would or would not appear on a given trial, remembered orientations could not be decoded based on visual cortex activation patterns, but decoding from parietal cortex was successful. However, when distractor presence was unpredictable, both visual and parietal cortex represented remembered orientations during both distractor-present and -absent trials. Several subsequent commentaries (Ester et al., 2016; Gayet et al., 2018; Scimeca et al., 2018; Postle and Yu, 2020; Lorenc and Sreenivasan, 2021) and empirical reports (Lorenc et al., 2018; Rademaker et al., 2019; Hallenbeck et al., in press) contested the theoretical importance of the null decoding performance in visual cortex for predictable distractors observed in Bettencourt and Xu (2016). The empirical studies largely replicated the finding in Bettencourt and Xu (2016) that activation patterns in parietal cortex contained information about WM content regardless of whether or not a distractor was present during the delay. However, decoded activation patterns in visual cortex do seem to depend on distractor presence, and alterations in these representations predict behavioral errors (Lorenc et al., 2018; Rademaker et al., 2019; Iamshchinina et al., 2021; Hallenbeck et al., in press). Thus, persistent activity, as indexed by successful decoding of remembered information, survives visual distraction in many regions, and the impact of distraction on measured persistent activity in visual cortex is reflected in behavioral performance errors. These results across several studies suggest a critical role for stimulus-selective persistent activity in sensory cortex–it is often observed during delay periods, it appears unaffected by distractors when behavioral performance remains intact, and changes in persistent activity are reflected in changes in behavioral responses.

CONCLUDING REMARKS: PERSISTENT ACTIVITY PERSISTS

Remarkably, for the past 50 years researchers studying the mechanisms of WM have used a variety of tools to characterize persistent activity across numerous types of memory tasks, across species, and across brain areas. It continues to stand as the central neural mechanism that supports WM. We have seen two arcs of research into persistent activity. One began in the front of the brain with single neuron recordings from the macaque PFC

and led to what we have referred to as a canonical model of WM–a rich and mechanistically explicit computational model based in physiology and anatomy. The other began in the back of the brain using sophisticated machine learning algorithms that could precisely decode the contents of WM based on the patterns of neural activity in the human visual system. Here, we argue that such decoding is itself a manifestation of stimulus-selective persistent activity, just at a smaller scale than entire brain regions. Accordingly, persistent activity can be inferred not just from sustained elevated spiking of neurons, but from population level activity of fMRI BOLD signals sculpted by the content of memory.

The recent challenges to the canonical model include the various coding schemes (e.g., dynamic coding, mixed selectivity) and concerns about the evidence for persistent activity itself (e.g., artifacts, activity silent mechanisms). Another implicit challenge revolves around how "PFC-centric" the field has been when considering the neural mechanisms of WM. For instance, even if we accept each of these criticisms, the canonical model in its simplest form would need *no* revision if it was merely applied to brain areas other than the PFC. For instance, translating the classic findings of Funahashi–firing of neurons in macaque dorsolateral PFC persist over delays (Funahashi et al., 1989) and damage to this region impacts WM (Funahashi et al., 1993a)–to humans only requires shifting the locus from dorsolateral PFC to a brain region a bit more posterior in the precentral/arcuate sulcus (Mackey et al., 2016b). The various complexities with the types of coding and reliability of persistent activity in monkey PFC all disappear if the canonical model is instead applied to monkey areas like FEF and LIP (Hart and Huk, 2020). In those areas, single neurons clearly persist on single trials and the neurons form populations that represent WM features exactly as modeled (Wang, 2001) without the complexity of mixed selectivity and dynamic coding. Perhaps while we attempt to reconcile new discoveries about the PFC, we do not need to update our canonical model of WM. Overwhelmingly, the evidence indicates that simple, stable persistent activity among neurons in stimulus selective populations is one fundamental mechanism by which we maintain WM representations.

Moving forward, there are a number of questions that we instead need to address. Most relevant is: what, then, *is* the role of the dorsolateral PFC? Note that persistent activity is simply an observation and is not synonymous with WM maintenance (Curtis and Lee, 2010). Perhaps persistent activity in PFC reflects not the storage of WM features, but rather some mechanism related to the control of WM representations stored elsewhere, maybe by their own persistent activity (Miller and Cohen, 2001;

Curtis and D'Esposito, 2003; Emrich et al., 2013; D'Esposito and Postle, 2015; Postle and Yu, 2020). Also, why do we see evidence of persistent activity, even for a simple single item WM task, in so many cortical and subcortical brain areas (Christophel et al., 2017; Leavitt et al., 2017)? Redundancy is good to a point, but future research should try to figure out which of these numerous areas are necessary, what types of features they might be representing, and if they might be encoding different representational formats of WM. For example, disrupting persistent activity with intervening distraction (Lorenc et al., 2018; Rademaker et al., 2019; e.g., Hallenbeck et al., in press) or TMS (e.g., Mackey and Curtis, 2017; Rademaker et al., 2017) may be able to disentangle the relative roles of different cortical regions. However, such efforts are tricky, as a distractor may not affect a top-down control signal, especially when passively viewed. Experiments parametrically manipulating task demands in concert with visual distraction may help further clarify the relative role different brain regions play in WM tasks. An especially promising avenue for future exploration is comparing decoded feature values from single trials of fMRI activation to behavioral errors on those same trials (Ester et al., 2016; Hallenbeck et al., in press; Li et al., in press).

Working memory is one of the few higher-level cognitive systems that we have made substantial progress toward understanding its neural implementation. Persistent activity has been at the heart of this success. While it is inevitable that additional mechanisms will be discovered, we have little doubt that persistent activity will persist as a primary explanation for how neural systems maintain WM representations. Future empirical research should focus on understanding the degree to which mechanisms are shared between the canonical and sensory recruitment models of WM, and the degree to which the challenges we highlighted in this review require revising the theoretical mechanisms that support WM.

AUTHOR CONTRIBUTIONS

Both authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

REFERENCES

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., and de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Curr. Biol.* 23, 1427–1431. doi: 10.1016/j.cub.2013.05.065

Alink, A., Krugliak, A., Walther, A., and Kriegeskorte, N. (2013). fMRI orientation decoding in V1 does not require global maps or globally coherent orientation stimuli. *Front. Psychol.* 4:493. doi: 10.3389/fpsyg.2013.00493

Allefeld, C., and Haynes, J.-D. (2014). Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA. *Neuroimage* 89, 345–357. doi: 10.1016/j.neuroimage.2013.11.043

Asaad, W. F., Rainer, G., and Miller, E. K. (2000). Task-specific neural activity in the primate prefrontal cortex. *J. Neurophysiol.* 84, 451–459. doi: 10.1152/jn.2000.84.1.451

Baddeley, A. D. (1986). *Working Memory*. Oxford: Oxford University Press.

Barash, S., Bracewell, R. M., Fogassi, L., Gnadt, J. W., and Andersen, R. A. (1991). Saccade-related activity in the lateral intraparietal area. II.

Spatial properties. *J. Neurophysiol.* 66, 1109–1124. doi: 10.1152/jn.1991.66.3.1109

Barbosa, J., Soldevilla, D. L., and Compte, A. (2021). Unattended short-term memories are maintained in active neural representations. *PsyArXiv* [Preprint]. doi: 10.31234/osf.io/qv6fu

Bastos, A. M., Loonis, R., Kornblith, S., Lundqvist, M., and Miller, E. K. (2018). Laminar recordings in frontal cortex suggest distinct layers for maintenance and control of working memory. *Proc. Natl. Acad. Sci. U.S.A.* 115, 1117–1122. doi: 10.1073/pnas.1710323115

Bettencourt, K. C., and Xu, Y. (2016). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nat. Neurosci.* 19, 150–157. doi: 10.1038/nn.4174

Bisley, J. W., Zaksas, D., Droll, J. A., and Pasternak, T. (2004). Activity of neurons in cortical area MT during a memory for motion task. *J. Neurophysiol.* 91, 286–300. doi: 10.1152/jn.00870.2003

Blatt, G. J., Andersen, R. A., and Stoner, G. R. (1990). Visual receptive field organization and cortico-cortical connections of the lateral intraparietal area (area LIP) in the macaque. *J. Comp. Neurol.* 299, 421–445. doi: 10.1002/cne.902990404

Boynton, G. M. (2005). Imaging orientation selectivity: decoding conscious perception in V1. *Nat. Neurosci.* 8, 541–542. doi: 10.1038/nn0505-541

Brissenden, J. A., Tobyne, S. M., Halko, M. A., and Somers, D. C. (2021). Stimulus-specific visual working memory representations in human cerebellar lobule VIIb/VIIIa. *J. Neurosci.* 41, 1033–1045. doi: 10.1523/jneurosci.1253-20.2020

Brouwer, G. J., and Heeger, D. J. (2009). Decoding and reconstructing color from responses in human visual cortex. *J. Neurosci.* 29, 13992–14003. doi: 10.1523/jneurosci.3577-09.2009

Carlson, T. A. (2014). Orientation decoding in human visual cortex: new insights from an unbiased perspective. *J. Neurosci.* 34, 8373–8383. doi: 10.1523/jneurosci.0548-14.2014

Cavanagh, S. E., Towers, J. P., Wallis, J. D., Hunt, L. T., and Kennerley, S. W. (2018). Reconciling persistent and dynamic hypotheses of working memory coding in prefrontal cortex. *Nat. Commun.* 9:3498.

Chafee, M. V., and Goldman-Rakic, P. S. (1998). Matching patterns of activity in primate prefrontal area 8a and Parietal Area 7ip neurons during a spatial working MemoryTask. *J. Neurophysiol.* 79, 2919–2940. doi: 10.1152/jn.1998.79.6.2919

Chelazzi, L., Duncan, J., Miller, E. K., and Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *J. Neurophysiol.* 80, 2918–2940. doi: 10.1152/jn.1998.80.6.2918

Christophel, T. B., Allefeld, C., Endisch, C., and Haynes, J.-D. (2018a). View-independent working memory representations of artificial shapes in prefrontal and posterior regions of the human brain. *Cereb. Cortex* 28, 2146–2161. doi: 10.1093/cercor/bhx119

Christophel, T. B., Iamshchinina, P., Yan, C., Allefeld, C., and Haynes, J.-D. (2018b). Cortical specialization for attended versus unattended working memory. *Nat. Neurosci.* 21, 494–496. doi: 10.1038/s41593-018-0094-4

Christophel, T. B., and Haynes, J.-D. (2014). Decoding complex flow-field patterns in visual working memory. *Neuroimage* 91, 43–51. doi: 10.1016/j.neuroimage.2014.01.025

Christophel, T. B., Hebart, M. N., and Haynes, J.-D. (2012). Decoding the contents of visual short-term memory from human visual and parietal cortex. *J. Neurosci.* 32, 12983–12989. doi: 10.1523/jneurosci.0184-12.2012

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., and Haynes, J.-D. (2017). the distributed nature of working memory. *Trends Cogn. Sci.* 21, 111–124.

Compte, A., Brunel, N., Goldman-Rakic, P. S., and Wang, X. J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex* 10, 910–923. doi: 10.1093/cercor/10.9.910

Constantinidis, C., Franowicz, M. N., and Goldman-Rakic, P. S. (2001). Coding specificity in cortical microcircuits: a multiple-electrode analysis of primate prefrontal cortex. *J. Neurosci.* 21, 3646–3655. doi: 10.1523/jneurosci.21-10-03646.2001

Constantinidis, C., and Goldman-Rakic, P. S. (2002). Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. *J. Neurophysiol.* 88, 3487–3497. doi: 10.1152/jn.00188.2002

Constantinidis, C., and Steinmetz, M. A. (1996). Neuronal activity in posterior parietal area 7a during the delay periods of a spatial memory task. *J. Neurophysiol.* 76, 1352–1355. doi: 10.1152/jn.1996.76.2.1352

Courtney, S. M., Petit, L., Maisog, J. M., Ungerleider, L. G., and Haxby, J. V. (1998). An area specialized for spatial working memory in human frontal cortex. *Science* 279, 1347–1351. doi: 10.1126/science.279.5355.1347

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* 24, 87–114. discussion 114–85. doi: 10.1017/s0140525x01003922

Curtis, C. E., and D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.* 7, 415–423. doi: 10.1016/s1364-6613(03)00197-9

Curtis, C. E., and D'Esposito, M. (2004). The effects of prefrontal lesions on working memory performance and theory. *Cogn. Affect. Behav. Neurosci.* 4, 528–539. doi: 10.3758/cabn.4.4.528

Curtis, C. E., and D'Esposito, M. (2006). Selection and maintenance of saccade goals in the human frontal eye fields. *J. Neurophysiol.* 95, 3923–3927. doi: 10.1152/jn.01120.2005

Curtis, C. E., and Lee, D. (2010). Beyond working memory: the role of persistent activity in decision making. *Trends Cogn. Sci.* 14, 216–222. doi: 10.1016/j.tics.2010.03.006

Curtis, C. E., Rao, V. Y., and D'Esposito, M. (2004). Maintenance of spatial and motor codes during oculomotor delayed response tasks. *J. Neurosci.* 24, 3944–3952. doi: 10.1523/jneurosci.5640-03.2004

Dash, S., Yan, X., Wang, H., and Crawford, J. D. (2015). Continuous updating of visuospatial memory in superior colliculus during slow eye movements. *Curr. Biol.* 25, 267–274. doi: 10.1016/j.cub.2014.11.064

D'Esposito, M., and Postle, B. R. (1999). The dependence of span and delayed-response performance on prefrontal cortex. *Neuropsychologia* 37, 1303–1315. doi: 10.1016/s0028-3932(99)00021-4

D'Esposito, M., and Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annu. Rev. Psychol.* 66, 115–142.

di Pellegrino, G., and Wise, S. P. (1993). Visuospatial versus visuomotor activity in the premotor and prefrontal cortex of a primate. *J. Neurosci.* 13, 1227–1243. doi: 10.1523/jneurosci.13-03-01227.1993

Dias, E. C., and Segraves, M. A. (1999). Muscimol-induced inactivation of monkey frontal eye field: effects on visually and memory-guided saccades. *J. Neurophysiol.* 81, 2191–2214. doi: 10.1152/jn.1999.81.5.2191

Dumoulin, S. O., and Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage* 39, 647–660. doi: 10.1016/j.neuroimage.2007.09.034

Emrich, S. M., Riggall, A. C., Larocque, J. J., and Postle, B. R. (2013). Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. *J. Neurosci.* 33, 6516–6523. doi: 10.1523/jneurosci.5732-12.2013

Ester, E. F., Anderson, D. E., Serences, J. T., and Awh, E. (2013). A neural measure of precision in visual working memory. *J. Cogn. Neurosci.* 25, 754–761. doi: 10.1162/jocn_a_00357

Ester, E. F., Rademaker, R. L., and Sprague, T. C. (2016). How do visual and parietal cortex contribute to visual short-term memory? *eNeuro* 3, ENEURO.41-ENEURO.2016. doi: 10.1523/ENEURO.0041-16.2016

Ester, E. F., Serences, J. T., and Awh, E. (2009). Spatially global representations in human primary visual cortex during working memory maintenance. *J. Neurosci.* 29, 15258–15265. doi: 10.1523/jneurosci.4388-09.2009

Ester, E. F., Sprague, T. C., and Serences, J. T. (2015). Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. *Neuron* 87, 893–905. doi: 10.1016/j.neuron.2015.07.013

Fecteau, J. H., and Munoz, D. P. (2006). Salience, relevance, and firing: a priority map for target selection. *Trends Cogn. Sci.* 10, 382–390. doi: 10.1016/j.tics.2006.06.011

Foster, J. J., Sutterer, D. W., Serences, J. T., Vogel, E. K., and Awh, E. (2016). The topography of alpha-band activity tracks the content of spatial working memory. *J. Neurophysiol.* 115, 168–177. doi: 10.1152/jn.00860.2015

Freedman, D. J., Riesenhuber, M., Poggio, T., and Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291, 312–316. doi: 10.1126/science.291.5502.312

Freeman, J., Brouwer, G. J., Heeger, D. J., and Merriam, E. P. (2011). Orientation decoding depends on maps, not columns. *J. Neurosci.* 31, 4792–4804. doi: 10.1523/jneurosci.5160-10.2011

Freeman, J., Heeger, D. J., and Merriam, E. P. (2013). Coarse-scale biases for spirals and orientation in human visual cortex. *J. Neurosci.* 33, 19695–19703. doi: 10.1523/jneurosci.0889-13.2013

Fukushima, J., Akao, T., Shichinohe, N., Kurkin, S., Kaneko, C. R. S., and Fukushima, K. (2011). Neuronal activity in the caudal frontal eye fields of monkeys during memory-based smooth pursuit eye movements: comparison with the supplementary eye fields. *Cereb. Cortex* 21, 1910–1924. doi: 10.1093/cercor/bhq261

Funahashi, S. (2013). Thalamic mediodorsal nucleus and its participation in spatial working memory processes: comparison with the prefrontal cortex. *Front. Syst. Neurosci.* 7:36. doi: 10.3389/fnsys.2013.00036

Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* 61, 331–349. doi: 10.1152/jn.1989.61.2.331

Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. (1993a). Dorsolateral prefrontal lesions and oculomotor delayed-response performance: evidence for mnemonic "scotomas.". *J. Neurosci.* 13, 1479–1497. doi: 10.1523/jneurosci.13-04-01479.1993

Funahashi, S., Chafee, M. V., and Goldman-Rakic, P. S. (1993b). Prefrontal neuronal activity in rhesus monkeys performing a delayed anti-saccade task. *Nature* 365, 753–756. doi: 10.1038/365753a0

Fuster, J. M., and Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science* 173, 652–654. doi: 10.1126/science.173.3997.652

Fuster, J. M., Bodner, M., and Kroger, J. K. (2000). Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature* 405, 347–351. doi: 10.1038/35012613

Fuster, J. M., and Jervey, J. P. (1981). Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science* 212, 952–955. doi: 10.1126/science.7233192

Gayet, S., Paffen, C. L. E., and Van der Stigchel, S. (2018). Visual working memory storage recruits sensory processing areas. *Trends Cogn. Sci.* 22, 189–190. doi: 10.1016/j.tics.2017.09.011

Gaymard, B., Ploner, C. J., Rivaud-Péchoux, S., and Pierrot-Deseilligny, C. (1999). The frontal eye field is involved in spatial short-term memory but not in reflexive saccade inhibition. *Exp. Brain Res.* 129, 288–301. doi: 10.1007/s002210050899

Gilbert, C. D., and Wiesel, T. N. (1983). Clustered intrinsic connections in cat visual cortex. *J. Neurosci.* 3, 1116–1133. doi: 10.1523/jneurosci.03-05-01116.1983

Gilbert, C. D., and Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *J. Neurosci.* 9, 2432–2442. doi: 10.1523/jneurosci.09-07-02432.1989

Gnadt, J. W., and Andersen, R. A. (1988). Memory related motor planning activity in posterior parietal cortex of macaque. *Exp. Brain Res.* 70, 216–220.

Goard, M. J., Pho, G. N., Woodson, J., and Sur, M. (2016). Distinct roles of visual, parietal, and frontal motor cortices in memory-guided sensorimotor decisions. *eLife* 5:e13764.

Goldberg, M. E., and Bruce, C. J. (1985). Cerebral cortical activity associated with the orientation of visual attention in the rhesus monkey. *Vision Res.* 25, 471–481. doi: 10.1016/0042-6989(85)90072-0

Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron* 14, 477–485. doi: 10.1016/0896-6273(95)90304-6

Hallenbeck, G. E., Sprague, T. C., Rahmati, M., Sreenivasan, K. K., and Curtis, C. E. (in press). Working memory representations in visual cortex mediate the effects of distraction. *Nat. Commun.*

Hamed, S. B., Ben Hamed, S., Duhamel, J.-R., Bremmer, F., and Graf, W. (2001). Representation of the visual field in the lateral intraparietal area of macaque monkeys: a quantitative receptive field analysis. *Exp. Brain Res.* 140, 127–144. doi: 10.1007/s002210100785

Harrison, S. A., and Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458, 632–635. doi: 10.1038/nature07832

Hart, E., and Huk, A. C. (2020). Recurrent circuit dynamics underlie persistent activity in the macaque frontoparietal network. *eLife* 9:e52460. doi: 10.7554/eLife.52460

Haxby, J. V., Ida Gobbini, M., Furey, M. L., Ishai, A., Schouten, J. L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425–2430. doi: 10.1126/science.1063736

Hayden, B. Y., and Gallant, J. L. (2013). Working memory and decision processes in visual area v4. *Front. Neurosci.* 7:18. doi: 10.3389/fnins.2013.00018

Haynes, J.-D., and Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat. Neurosci.* 8, 686–691. doi: 10.1038/nn1445

Hwang, J., and Romanski, L. M. (2015). Prefrontal neuronal responses during audiovisual mnemonic processing. *J. Neurosci.* 35, 960–971. doi: 10.1523/jneurosci.1328-14.2015

Iamshchinina, P., Christophel, T. B., Gayet, S., and Rademaker, R. L. (2021). Essential considerations for exploring visual working memory storage in the human brain. *Vis. Cogn.* 29, 1–12. doi: 10.1080/13506285.2021.1915902

Ichihara-Takeda, S., and Funahashi, S. (2007). Activity of primate orbitofrontal and dorsolateral prefrontal neurons: task-related activity during an oculomotor delayed-response task. *Exp. Brain Res.* 181, 409–425. doi: 10.1007/s00221-007-0941-0

Inagaki, H. K., Fontolan, L., Romani, S., and Svoboda, K. (2019). Discrete attractor dynamics underlies persistent activity in the frontal cortex. *Nature* 566, 212–217. doi: 10.1038/s41586-019-0919-7

Jerde, T. A., Merriam, E. P., Riggall, A. C., Hedges, J. H., and Curtis, C. E. (2012). Prioritized maps of space in human frontoparietal cortex. *J. Neurosci.* 32, 17382–17390. doi: 10.1523/jneurosci.3810-12.2012

Jonides, J., Smith, E. E., Koeppe, R. A., Awh, E., Minoshima, S., and Mintun, M. A. (1993). Spatial working memory in humans as revealed by PET. *Nature* 363, 623–625. doi: 10.1038/363623a0

Kamiński, J., Sullivan, S., Chung, J. M., Ross, I. B., Mamelak, A. N., and Rutishauser, U. (2017). Persistently active neurons in human medial frontal and medial temporal lobe support working memory. *Nat. Neurosci.* 20, 590–601. doi: 10.1038/nn.4509

Kamitani, Y., and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* 8, 679–685. doi: 10.1038/nn1444

Kikuchi-Yorioka, Y., and Sawaguchi, T. (2000). Parallel visuospatial and audiospatial working memory processes in the monkey dorsolateral prefrontal cortex. *Nat. Neurosci.* 3, 1075–1076. doi: 10.1038/80581

Kritzer, M. F., and Goldman-Rakic, P. S. (1995). Intrinsic circuit organization of the major layers and sublayers of the dorsolateral prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.* 359, 131–143. doi: 10.1002/cne.903590109

Kubota, K., and Niki, H. (1971). Prefrontal cortical unit activity and delayed alternation performance in monkeys. *J. Neurophysiol.* 34, 337–347. doi: 10.1152/jn.1971.34.3.337

LaRocque, J. J., Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., and Postle, B. R. (2013). Decoding attended information in short-term memory: an EEG study. *J. Cogn. Neurosci.* 25, 127–142. doi: 10.1162/jocn_a_00305

LaRocque, J. J., Riggall, A. C., Emrich, S. M., and Postle, B. R. (2017). Within-category decoding of information in different attentional states in short-term memory. *Cereb. Cortex* 27, 4881–4890.

Lawrence, S. J. D., van Mourik, T., Kok, P., Koopmans, P. J., Norris, D. G., and de Lange, F. P. (2018). Laminar organization of working memory signals in human visual cortex. *Curr. Biol.* 28, 3435.e4–3440.e4.

Leavitt, M. L., Mendoza-Halliday, D., and Martinez-Trujillo, J. C. (2017). sustained activity encoding working memories: not fully distributed. *Trends Neurosci.* 40, 328–346. doi: 10.1016/j.tins.2017.04.004

Leung, H.-C., Gore, J. C., and Goldman-Rakic, P. S. (2002). Sustained mnemonic response in the human middle frontal gyrus during on-line storage of spatial memoranda. *J. Cogn. Neurosci.* 14, 659–671. doi: 10.1162/08989290260045882

Levitt, J. B., Lewis, D. A., Yoshioka, T., and Lund, J. S. (1993). Topography of pyramidal neuron intrinsic connections in macaque monkey prefrontal cortex (areas 9 and 46). *J. Comp. Neurol.* 338, 360–376. doi: 10.1002/cne.903380304

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

Li, C.-S. R., Mazzoni, P., and Andersen, R. A. (1999). Effect of reversible inactivation of macaque lateral intraparietal area on visual and memory saccades. *J. Neurophysiol.* 81, 1827–1838. doi: 10.1152/jn.1999.81.4.1827

Li, H.-H., Sprague, T. C., Yoo, A. H., Ma, W. J., and Curtis, C. E. (in press). Joint representation of working memory and uncertainty in human cortex. *Neuron.*

Logie, R. H., Zucco, G. M., and Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta Psychol.* 75, 55–74. doi: 10.1016/0001-6918(90) 90066-o

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

Logothetis, N. K., and Wandell, B. A. (2004). Interpreting the BOLD signal. *Annu. Rev. Physiol.* 66, 735–769. doi: 10.1146/annurev.physiol.66.082602.092845

Lorenc, E. S., Sreenivasan, K. K., Nee, D. E., Vandenbroucke, A. R. E., and D'Esposito, M. (2018). Flexible coding of visual working memory representations during distraction. *J. Neurosci.* 38, 5267–5276. doi: 10.1523/jneurosci.3061-17.2018

Lorenc, E. S., Vandenbroucke, A. R. E., Nee, D. E., de Lange, F. P., and D'Esposito, M. (2020). Dissociable neural mechanisms underlie currently-relevant, future-relevant, and discarded working memory representations. *Sci. Rep.* 10:11195.

Lorenc, E. S., and Sreenivasan, K. K. (2021). Reframing the debate: the distributed systems view of working memory. *Vis. Cogn.* 29, 1–9. doi: 10.1080/13506285. 2021.1899091

Lund, J. S., Yoshioka, T., and Levitt, J. B. (1993). Comparison of intrinsic connectivity in different areas of macaque monkey cerebral cortex. *Cereb. Cortex* 3, 148–162. doi: 10.1093/cercor/3.2.148

Lundqvist, M., Herman, P., Warden, M. R., Brincat, S. L., and Miller, E. K. (2018). Gamma and beta bursts during working memory readout suggest roles in its volitional control. *Nat. Commun.* 9:394.

Lundqvist, M., Rose, J., Herman, P., Brincat, S. L., Buschman, T. J., and Miller, E. K. (2016). Gamma and beta bursts underlie working memory. *Neuron* 90, 152–164. doi: 10.1016/j.neuron.2016.02.028

Machens, C. K. (2010). Demixing population activity in higher cortical areas. *Front. Comput. Neurosci.* 4:126. doi: 10.3389/fncom.2010.00126

Machens, C. K., Romo, R., and Brody, C. D. (2010). Functional, but not anatomical, separation of "what" and "when" in prefrontal cortex. *J. Neurosci.* 30, 350–360. doi: 10.1523/jneurosci.3276-09.2010

Mackey, W. E., and Curtis, C. E. (2017). Distinct contributions by frontal and parietal cortices support working memory. *Sci. Rep.* 7:6188.

Mackey, W. E., Devinsky, O., Doyle, W. K., Golfinos, J. G., and Curtis, C. E. (2016a). Human parietal cortex lesions impact the precision of spatial working memory. *J. Neurophysiol.* 116, 1049–1054. doi: 10.1152/jn.00380.2016

Mackey, W. E., Devinsky, O., Doyle, W. K., Meager, M. R., and Curtis, C. E. (2016b). Human dorsolateral prefrontal cortex is not necessary for spatial working memory. *J. Neurosci.* 36, 2847–2856. doi: 10.1523/jneurosci.3618-15. 2016

Mackey, W. E., Winawer, J., and Curtis, C. E. (2017). Visual field map clusters in human frontoparietal cortex. *eLife* 6:e22974. doi: 10.7554/eLife.22974

Maloney, R. T. (2015). The basis of orientation decoding in human primary visual cortex: fine- or coarse-scale biases? *J. Neurophysiol.* 113, 1–3. doi: 10.1152/jn. 00196.2014

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

Markowitz, D. A., Curtis, C. E., and Pesaran, B. (2015). Multiple component networks support working memory in prefrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 112, 11084–11089. doi: 10.1073/pnas.1504172112

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102, 419–457. doi: 10.1037/0033-295x.102. 3.419

Mendoza-Halliday, D., Torres, S., and Martinez-Trujillo, J. C. (2014). Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nat. Neurosci.* 17, 1255–1262. doi: 10.1038/nn.3785

Merrikhi, Y., Clark, K., Albarran, E., Parsa, M., Zirnsak, M., Moore, T., et al. (2017). Spatial working memory alters the efficacy of input to visual cortex. *Nat. Commun.* 8:15041.

Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro.24.1. 167

Miller, E. K., and Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science* 263, 520–522. doi: 10.1126/science.8290960

Miller, E. K., Erickson, C. A., and Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J. Neurosci.* 16, 5154–5167. doi: 10.1523/jneurosci.16-16-05154.1996

Miller, E. K., Li, L., and Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *J. Neurosci.* 13, 1460–1478. doi: 10.1523/jneurosci.13-04-01460.1993

Miyashita, Y., and Chang, H. S. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature* 331, 68–70. doi: 10.1038/ 331068a0

Mohler, C. W., Goldberg, M. E., and Wurtz, R. H. (1973). Visual receptive fields of frontal eye field neurons. *Brain Res.* 61, 385–389. doi: 10.1016/0006-8993(73) 90543-x

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

Murray, J. D., Bernacchia, A., Roy, N. A., Constantinidis, C., Romo, R., and Wang, X.-J. (2017). Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 114, 394–399. doi: 10.1073/pnas.1619449114

Norman, K. A., Polyn, S. M., Detre, G. J., and Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* 10, 424–430. doi: 10.1016/j.tics.2006.07.005

Ó Scalaidhe, S. P., Wilson, F. A. W., and Goldman-Rakic, P. S. (1999). Face-selective neurons during passive viewing and working memory performance of rhesus monkeys: evidence for intrinsic specialization of neuronal coding. *Cereb. Cortex* 9, 459–475. doi: 10.1093/cercor/9.5.459

Offen, S., Schluppeck, D., and Heeger, D. J. (2009). The role of early visual cortex in visual short-term memory and visual attention. *Vis. Res.* 49, 1352–1362. doi: 10.1016/j.visres.2007.12.022

Oga, T., Elston, G. N., and Fujita, I. (2017). Postnatal dendritic growth and spinogenesis of layer-v pyramidal cells differ between visual, inferotemporal, and prefrontal cortex of the macaque monkey. *Front. Neurosci.* 11:118. doi: 10.3389/fnins.2017.00118

Panichello, M. F., and Buschman, T. J. (2021). Shared mechanisms underlie the control of working memory and attention. *Nature* 592, 601–605. doi: 10.1038/ s41586-021-03390-w

Parthasarathy, A., Herikstad, R., Bong, J. H., Medina, F. S., Libedinsky, C., and Yen, S.-C. (2017). Mixed selectivity morphs population codes in prefrontal cortex. *Nat. Neurosci.* 20, 1770–1779. doi: 10.1038/s41593-017-0003-2

Parthasarathy, A., Tang, C., Herikstad, R., Cheong, L. F., Yen, S.-C., and Libedinsky, C. (2019). Time-invariant working memory representations in the presence of code-morphing in the lateral prefrontal cortex. *Nat. Commun.* 10:4995.

Pasternak, T., and Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nat. Rev. Neurosci.* 6, 97–107. doi: 10.1038/nrn1603

Pesaran, B., Pezaris, J. S., Sahani, M., Mitra, P. P., and Andersen, R. A. (2002). Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat. Neurosci.* 5, 805–811. doi: 10.1038/nn890

Petrides, M., Tomaiuolo, F., Yeterian, E. H., and Pandya, D. N. (2012). The prefrontal cortex: comparative architectonic organization in the human and the macaque monkey brains. *Cortex* 48, 46–57. doi: 10.1016/j.cortex.2011.07.002

Ploner, C. J., Rivaud-Péchoux, S., Gaymard, B. M., Agid, Y., and Pierrot-Deseilligny, C. (1999). Errors of memory-guided saccades in humans with lesions of the frontal eye field and the dorsolateral prefrontal cortex. *J. Neurophysiol.* 82, 1086–1090. doi: 10.1152/jn.1999.82.2.1086

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience* 139, 23–38. doi: 10.1016/j.neuroscience.2005.06.005

Postle, B. R. (2015). The cognitive neuroscience of visual short-term memory. *Curr. Opin. Behav. Sci.* 1, 40–46. doi: 10.1016/j.cobeha.2014.08.004

Postle, B. R., Druzgal, T. J., and D'Esposito, M. (2003). Seeking the neural substrates of visual working memory storage. *Cortex* 39, 927–946. doi: 10.1016/s0010-9452(08)70871-2

Postle, B. R., Idzikowski, C., Sala, S. D., Logie, R. H., and Baddeley, A. D. (2006). The selective disruption of spatial working memory by eye movements. *Q. J. Exp. Psychol.* 59, 100–120. doi: 10.1080/17470210500 151410

Postle, B. R., and Yu, Q. (2020). Neuroimaging and the localization of function in visual cognition. *Vis. Cogn.* 28, 447–452. doi: 10.1080/13506285.2020.1777237

Pratte, M. S., Sy, J. L., Swisher, J. D., and Tong, F. (2016). Radial bias is not necessary for orientation decoding. *Neuroimage* 127, 23–33. doi: 10.1016/j.neuroimage.2015.11.066

Quintana, J., Yajeya, J., and Fuster, J. M. (1988). Prefrontal representation of stimulus attributes during delay tasks. I. Unit activity in cross-temporal integration of sensory and sensory-motor information. *Brain Res.* 474, 211–221. doi: 10.1016/0006-8993(88)90436-2

Rademaker, R. L., Chunharas, C., and Serences, J. T. (2019). Coexisting representations of sensory and mnemonic information in human visual cortex. *Nat. Neurosci.* 22, 1336–1344. doi: 10.1038/s41593-019-0428-x

Rademaker, R. L., van de Ven, V. G., Tong, F., and Sack, A. T. (2017). The impact of early visual cortex transcranial magnetic stimulation on visual working memory precision and guess rate. *PLoS One* 12:e0175230. doi: 10.1371/journal.pone.0175230

Rahmati, M., DeSimone, K., Curtis, C. E., and Sreenivasan, K. K. (2020). Spatially specific working memory activity in the human superior colliculus. *J. Neurosci.* 40, 9487–9495. doi: 10.1523/jneurosci.2016-20.2020

Rahmati, M., Saber, G. T., and Curtis, C. E. (2018). Population dynamics of early visual cortex during working memory. *J. Cogn. Neurosci.* 30, 219–233. doi: 10.1162/jocn_a_01196

Rainer, G., Asaad, W. F., and Miller, E. K. (1998). Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature* 393, 577–579. doi: 10.1038/31235

Rainer, G., Rao, S. C., and Miller, E. K. (1999). Prospective coding for objects in primate prefrontal cortex. *J. Neurosci.* 19, 5493–5505. doi: 10.1523/jneurosci.19-13-05493.1999

Rao, S. C., Rainer, G., and Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science* 276, 821–824. doi: 10.1126/science.276.5313.821

Rao, S. G., Williams, G. V., and Goldman-Rakic, P. S. (2000). Destruction and creation of spatial tuning by disinhibition: GABAABlockade of prefrontal cortical neurons engaged by working memory. *J. Neurosci.* 20, 485–494. doi: 10.1523/jneurosci.20-01-00485.2000

Rauch, A., Rainer, G., and Logothetis, N. K. (2008). The effect of a serotonin-induced dissociation between spiking and perisynaptic activity on BOLD functional MRI. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6759–6764. doi: 10.1073/pnas.0800312105

Riggall, A. C., and Postle, B. R. (2012). The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *J. Neurosci.* 32, 12990–12998. doi: 10.1523/jneurosci.1892-12.2012

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

Riley, M. R., and Constantinidis, C. (2015). Role of prefrontal persistent activity in working memory. *Front. Syst. Neurosci.* 9:181. doi: 10.3389/fnsys.2015.00181

Rolls, E. T., and Treves, A. (1990). The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. *Network* 1, 407–421. doi: 10.1088/0954-898x_1_4_002

Romo, R., Brody, C. D., Hernández, A., and Lemus, L. (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* 399, 470–473. doi: 10.1038/20939

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

Rossi-Pool, R., Zainos, A., Alvarez, M., Zizumbo, J., Vergara, J., and Romo, R. (2017). Decoding a decision process in the neuronal population of dorsal premotor cortex. *Neuron* 96, 1432.e7–1446.e7.

Roth, Z. N., Heeger, D. J., and Merriam, E. P. (2018). Stimulus vignetting and orientation selectivity in human visual cortex. *eLife* 7:e37241. doi: 10.7554/eLife.37241

Rowe, J. B., Toni, I., Josephs, O., Frackowiak, R. S., and Passingham, R. E. (2000). The prefrontal cortex: response selection or maintenance within working memory? *Science* 288, 1656–1660. doi: 10.1126/science.288.5471.1656

Rypma, B., Berger, J. S., and D'Esposito, M. (2002). The influence of working-memory demand and subject performance on prefrontal cortical activity. *J. Cogn. Neurosci.* 14, 721–731. doi: 10.1162/08989290260138627

Saber, G. T., Pestilli, F., and Curtis, C. E. (2015). Saccade planning evokes topographically specific activity in the dorsal and ventral streams. *J. Neurosci.* 35, 245–252. doi: 10.1523/jneurosci.1687-14.2015

Sadeh, M., Sajad, A., Wang, H., Yan, X., and Crawford, J. D. (2018). The influence of a memory delay on spatial coding in the superior colliculus: is visual always visual and motor always motor? *Front. Neural Circuits* 12:74. doi: 10.3389/fncir.2018.00074

Schluppeck, D., Curtis, C. E., Glimcher, P. W., and Heeger, D. J. (2006). Sustained activity in topographic areas of human posterior parietal cortex during memory-guided saccades. *J. Neurosci.* 26, 5098–5108. doi: 10.1523/jneurosci.5330-05.2006

Schneegans, S., and Bays, P. M. (2017). Restoration of fMRI decodability does not imply latent working memory states. *J. Cogn. Neurosci.* 29, 1977–1994. doi: 10.1162/jocn_a_01180

Scimeca, J. M., Kiyonaga, A., and D'Esposito, M. (2018). Reaffirming the sensory recruitment account of working memory. *Trends Cogn. Sci.* 22, 190–192. doi: 10.1016/j.tics.2017.12.007

Serences, J. T. (2016). Neural mechanisms of information storage in visual short-term memory. *Vis. Res.* 128, 53–67. doi: 10.1016/j.visres.2016.09.010

Serences, J. T., Ester, E. F., Vogel, E. K., and Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychol. Sci.* 20, 207–214. doi: 10.1111/j.1467-9280.2009.02276.x

Serences, J. T., and Yantis, S. (2004). Attentional priority maps in human cortex. *PsycEXTRA Dataset* [Preprint]. doi: 10.1037/e537052012-668 [

Shafi, M., Zhou, Y., Quintana, J., Chow, C., Fuster, J., and Bodner, M. (2007). Variability in neuronal activity in primate cortex during working memory tasks. *Neuroscience* 146, 1082–1108. doi: 10.1016/j.neuroscience.2006.12.072

Shen, K., Valero, J., Day, G. S., and Paré, M. (2011). Investigating the role of the superior colliculus in active vision with the visual search paradigm. *Eur. J. Neurosci.* 33, 2003–2016. doi: 10.1111/j.1460-9568.2011.07722.x

Shichinohe, N., Akao, T., Kurkin, S., Fukushima, J., Kaneko, C. R. S., and Fukushima, K. (2009). Memory and decision making in the frontal cortex during visual motion processing for smooth pursuit eye movements. *Neuron* 62, 717–732. doi: 10.1016/j.neuron.2009.05.010

Sigala, N., Kusunoki, M., Nimmo-Smith, I., Gaffan, D., and Duncan, J. (2008). Hierarchical coding for sequential task events in the monkey prefrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11969–11974. doi: 10.1073/pnas.0802569105

Smith, E. E., Jonides, J., and Koeppe, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cereb. Cortex* 6, 11–20. doi: 10.1093/cercor/6.1.11

Sommer, M. A., and Wurtz, R. H. (2001). Frontal eye field sends delay activity related to movement, memory, and vision to the superior colliculus. *J. Neurophysiol.* 85, 1673–1685. doi: 10.1152/jn.2001.85.4.1673

Spaak, E., Watanabe, K., Funahashi, S., and Stokes, M. G. (2017). Stable and dynamic coding for working memory in primate prefrontal cortex. *J. Neurosci.* 37, 6503–6516. doi: 10.1523/jneurosci.3364-16.2017

Sprague, T. C., Adam, K. C. S., Foster, J. J., Rahmati, M., Sutterer, D. W., and Vo, V. A. (2018). Inverted encoding models assay population-level stimulus representations, not single-unit neural tuning. *eNeuro* 5, ENEURO.98-ENEURO.2018. doi: 10.1523/ENEURO.0098-18.2018

Sprague, T. C., Ester, E. F., and Serences, J. T. (2014). Reconstructions of information in visual spatial working memory degrade with memory load. *Curr. Biol.* 24, 2174–2180. doi: 10.1016/j.cub.2014.07.066

Sprague, T. C., Ester, E. F., and Serences, J. T. (2016). Restoring latent visual working memory representations in human cortex. *Neuron* 91, 694–707. doi: 10.1016/j.neuron.2016.07.006

Sreenivasan, K. K., Curtis, C. E., and D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends Cogn. Sci.* 18, 82–89. doi: 10.1016/j.tics.2013.12.001

Sreenivasan, K. K., and D'Esposito, M. (2019). The what, where and how of delay activity. *Nat. Rev. Neurosci.* 20, 466–481. doi: 10.1038/s41583-019-0176-7

Srimal, R., and Curtis, C. E. (2008). Persistent neural activity during the maintenance of spatial position in working memory. *Neuroimage* 39, 455–468. doi: 10.1016/j.neuroimage.2007.08.040

Standage, D., and Paré, M. (2018). Slot-like capacity and resource-like coding in a neural model of multiple-item working memory. *J. Neurophysiol.* 120, 1945–1961. doi: 10.1152/jn.00778.2017

Stokes, M., and Spaak, E. (2016). The importance of single-trial analyses in cognitive neuroscience. *Trends Cogn. Sci.* 20, 483–486. doi: 10.1016/j.tics.2016.05.008

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

Supèr, H., Spekreijse, H., and Lamme, V. A. (2001). A neural correlate of working memory in the monkey primary visual cortex. *Science* 293, 120–124. doi: 10.1126/science.1060496

Suzuki, M., and Gottlieb, J. (2013). Distinct neural mechanisms of distractor suppression in the frontal and parietal lobe. *Nat. Neurosci.* 16, 98–104. doi: 10.1038/nn.3282

Suzuki, W. A., Miller, E. K., and Desimone, R. (1997). Object and place memory in the macaque entorhinal cortex. *J. Neurophysiol.* 78, 1062–1081. doi: 10.1152/jn.1997.78.2.1062

Swisher, J. D., Gatenby, J. C., Gore, J. C., Wolfe, B. A., Moon, C.-H., Kim, S.-G., et al. (2010). Multiscale pattern analysis of orientation-selective activity in the primary visual cortex. *J. Neurosci.* 30, 325–330. doi: 10.1523/jneurosci.4811-09.2010

Takeda, K., and Funahashi, S. (2002). Prefrontal task-related activity representing visual cue location or saccade direction in spatial working memory tasks. *J. Neurophysiol.* 87, 567–588. doi: 10.1152/jn.00249.2001

Tark, K.-J., and Curtis, C. E. (2009). Persistent neural activity in the human frontal cortex when maintaining space that is off the map. *Nat. Neurosci.* 12, 1463–1468. doi: 10.1038/nn.2406

van Bergen, R. S., and Jehee, J. F. M. (2018). Modeling correlated noise is necessary to decode uncertainty. *Neuroimage* 180, 78–87. doi: 10.1016/j.neuroimage.2017.08.015

van Bergen, R. S., and Jehee, J. F. M. (2021). TAFKAP: an improved method for probabilistic decoding of cortical activity. *bioRxiv* [Preprint]. doi: 10.1101/2021.03.04.433946

van Bergen, R. S., Ma, W. J., Pratte, M. S., and Jehee, J. F. M. (2015). Sensory uncertainty decoded from visual cortex predicts behavior. *Nat. Neurosci.* 18, 1728–1730. doi: 10.1038/nn.4150

van Kerkoerle, T., Self, M. W., and Roelfsema, P. R. (2017). Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nat. Commun.* 8:13804.

Wallis, J. D., Anderson, K. C., and Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature* 411, 953–956. doi: 10.1038/35082081

Wandell, B. A., and Winawer, J. (2015). Computational neuroimaging and population receptive fields. *Trends Cogn. Sci.* 19, 349–357. doi: 10.1016/j.tics.2015.03.009

Wang, H., Stradtman, G. G., Wang, X.-J., and Gao, W.-J. (2008). A specialized NMDA receptor function in layer 5 recurrent microcircuitry of the adult rat prefrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16791–16796. doi: 10.1073/pnas.0804318105

Wang, M., Gamo, N. J., Yang, Y., Jin, L. E., Wang, X.-J., Laubach, M., et al. (2011). Neuronal basis of age-related working memory decline. *Nature* 476, 210–213. doi: 10.1038/nature10243

Wang, M., Yang, Y., Wang, C.-J., Gamo, N. J., Jin, L. E., Mazer, J. A., et al. (2013). NMDA receptors subserve persistent neuronal firing during working memory in dorsolateral prefrontal cortex. *Neuron* 77, 736–749. doi: 10.1016/j.neuron.2012.12.032

Wang, X. J. (1999). Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. *J. Neurosci.* 19, 9587–9603. doi: 10.1523/jneurosci.19-21-09587.1999

Wang, X. J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci.* 24, 455–463. doi: 10.1016/s0166-2236(00)01868-3

Wang, Y., Markram, H., Goodman, P. H., Berger, T. K., Ma, J., and Goldman-Rakic, P. S. (2006). Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nat. Neurosci.* 9, 534–542. doi: 10.1038/nn1670

Wasmuht, D. F., Spaak, E., Buschman, T. J., Miller, E. K., and Stokes, M. G. (2018). Intrinsic neuronal dynamics predict distinct functional roles during working memory. *Nat. Commun.* 9:3499.

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

Wirth, S., Yanike, M., Frank, L. M., Smith, A. C., Brown, E. N., and Suzuki, W. A. (2003). Single neurons in the monkey hippocampus and learning of new associations. *Science* 300, 1578–1581. doi: 10.1126/science.1084324

Wolff, M. J., Ding, J., Myers, N. E., and Stokes, M. G. (2015). Revealing hidden states in visual working memory using electroencephalography. *Front. Syst. Neurosci.* 9:123. doi: 10.3389/fnsys.2015.00123

Wolff, M. J., Jochim, J., Akyürek, E. G., Buschman, T. J., and Stokes, M. G. (2020a). Drifting codes within a stable coding scheme for working memory. *PLoS Biol.* 18:e3000625. doi: 10.1371/journal.pbio.3000625

Wolff, M. J., Kandemir, G., Stokes, M. G., and Akyürek, E. G. (2020b). Unimodal and bimodal access to sensory working memories by auditory and visual impulses. *J. Neurosci.* 40, 671–681. doi: 10.1523/jneurosci.1194-19.2019

Wolff, M. J., Jochim, J., Akyürek, E. G., and Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. *Nat. Neurosci.* 20, 864–871. doi: 10.1038/nn.4546

Yu, Q., and Shim, W. M. (2017). Occipital, parietal, and frontal cortices selectively maintain task-relevant features of multi-feature objects in visual working memory. *Neuroimage* 157, 97–107. doi: 10.1016/j.neuroimage.2017.05.055

Zaksas, D., and Pasternak, T. (2006). Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. *J. Neurosci.* 26, 11726–11742. doi: 10.1523/jneurosci.3420-06.2006

Zarahn, E., Aguirre, G. K., and D'Esposito, M. (1999). Temporal isolation of the neural correlates of spatial mnemonic processing with fMRI. *Brain Res. Cogn. Brain Res.* 7, 255–268. doi: 10.1016/s0926-6410(98)00029-9

Zelinsky, G. J., and Bisley, J. W. (2015). The what, where, and why of priority maps and their interactions with visual working memory. *Ann. N. Y. Acad. Sci.* 1339, 154–164. doi: 10.1111/nyas.12606