



HHS Public Access

Author manuscript

Compr Physiol. Author manuscript; available in PMC 2022 September 23.

Published in final edited form as:

Compr Physiol. ; 11(4): 2547–2587. doi:10.1002/cphy.c210005.

Working Memory: From Neural Activity to the Sentient Mind

Russell J. Jaffe¹, Christos Constantinidis^{1,2,3,4}

¹Department of Neurobiology & Anatomy, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA.

²Department of Biomedical Engineering, Vanderbilt University, Nashville TN 37235, USA.

³Neuroscience Program, Vanderbilt University, Nashville TN 37235, USA.

⁴Department of Ophthalmology and Visual Sciences, Vanderbilt University Medical Center, Nashville TN 37232, USA.

Abstract

Working memory is the ability to maintain and manipulate information in the conscious mind over a timescale of seconds. This ability is thought to be maintained through the persistent discharges of neurons in a network of brain areas centered on the prefrontal cortex, as evidenced by neurophysiological recordings in non-human primates, though both the localization and the neural basis of working memory has been a matter of debate in recent years. Neural correlates of working memory are evident in species other than primates, including rodents and corvids. A specialized network of excitatory and inhibitory neurons, aided by neuromodulatory influences of dopamine, is critical for the maintenance of neuronal activity. Limitations in working memory capacity and duration, as well as its enhancement during development can be attributed to properties of neural activity and circuits. Changes in these factors can be observed through training-induced improvements and in pathological impairments. Working memory thus provides a prototypical cognitive function whose properties can be tied to the spiking activity of brain neurons.

Introduction

Working memory (WM)—the ability to maintain and manipulate information in conscious mind over a timescale of seconds—is a critical cognitive function in the ability to learn, make decisions, and function in daily life (1, 2). This ability can be disrupted after brain injury, most importantly in the prefrontal cortex (PFC) (3-7). Neural correlates of WM have been identified in the activity of neurons in cortical and subcortical areas. As such, working memory represents a prototypical case of a mental phenomenon that can be explained in terms of underlying brain activity. The precise mechanisms that underlie WM remain subject to debate (8). In this article, we review the current state of knowledge and open questions, examining prior models and proposing a road forward.

We begin by reviewing conceptual models of WM as general frameworks that can be used to bind experimental findings into a unified theory. This discussion will also demonstrate how some of the controversies surrounding the field can be accounted for by the different methodologies and levels of analysis that different studies have undertaken. We proceed by reviewing the alternative accounts of the neural basis of WM, including those based on

persistent activity generated by cortical neurons, those based on rhythmic discharges and those that do not depend on discharge rates modulation (activity-silent models). The debate between these accounts also provides an opportunity to examine the localization of WM in the brain. Traditional models, such as the original bump attractor model, have placed the PFC as the seat of WM, though alternative accounts suggest that the sensory cortex is the primary site of WM, with the PFC playing a supervisory role instead.

Although the emphasis of the article is on visual WM, neural correlates of other modalities have also been described and their review is instructive to defining the underlying mechanisms of WM. Human WM is notoriously limited in terms of its capacity and duration; we examine the neural basis of these limitations, and the competing models that have been proposed to account for them. We then discuss neural correlates of WM in non-primates (rodent and avian species) and what they reveal about the evolution of working memory maintenance. We move on to examine the enhancement of working memory abilities through the course of childhood development and through training, in adulthood. We also discuss individual differences between people, the relationship between working memory and other constructs such as attention and intelligence, and disorders of working memory. The review ends with conclusions and open questions for future investigation.

Conceptual Models of Working Memory

The definition of working memory has undergone a series of revisions since its introduction to the scientific vernacular in the 1960's (9), better relating this mental phenomenon to well-defined neuroscience concepts in order to alleviate the indeterminacy that often relates to philosophical constructs (10, 11). WM is therefore conventionally defined at present as a fundamental cognitive system that facilitates the temporary storage and manipulation of information in the immediately conscious mind (1, 12-14). This is a critical cognitive function in the ability to learn, make decisions, and function in daily life, and its underlying mechanisms are often examined through the life disruptions that are suffered in the case of brain injuries or conditions such as schizophrenia and ADHD (3-7).

The classic Atkinson–Shiffrin multi-store model (Fig. 1A), also known as the modal model of memory, distinguished between three memory stores: a sensory register, a short-term store, and a long-term memory store (15). The sensory register buffers information from sensory modalities, and in the case of vision (“iconic” memory) it has very rapid decay, in the order of half a second or less. The duration of the sensory register for auditory information (“echoic” memory) is longer in the order of 5 seconds or so. As we will discuss in the later sections regarding the neural basis of WM, certain aspects of this model have a strong empirical grounding in current neuroscience research. The activation of neurons in the visual cortex persists for several hundred milliseconds after the offset of a stimulus, thus providing a neural correlate for the sensory register (16). The PFC and its connected areas then maintain specific patterns of activity throughout the entire period in which a stimulus is remembered, providing a neural correlate for the short-term store. Atkinson and Shiffrin also referred to the short-term store as working memory, using a term introduced by Miller and colleagues (9, 15, 17). Despite its appeal, the Atkinson–Shiffrin model was not without shortcomings. It assumed a single short-term memory system for all memory items and it

proposed rehearsal as the main mechanism for the transfer of information from short-term to long-term memory, without which, information would decay (18).

Spurred by experimental observations that contradicted these tenets, Baddeley and Hitch put forth a model of working memory (Fig. 1B), the most recent version of which distinguishes between a central executive and three subsidiary systems, the phonological loop, the visuospatial sketchpad, and episodic buffer (19). Auditory information, spoken and written words, and any other phonological information is thought to be stored in the phonological loop, which retains a limited store of objects through the process of real time vocal and subvocal rehearsal. The visual-spatial sketchpad plays a similar role for visual and spatial information, with limited interference from items maintained in the phonological loop. The most recent addition to the Baddeley model is the episodic buffer, which is thought to be a passive storage system that would link the different memory domains (1). This is a highly important function, as the integration of objects from different types of WM would allow the creation of entirely new objects from our imagination. Moreover, given the limits of working memory, the episodic buffer would serve the critical purpose of allowing a collection of associated features to be stored as a single complex object, to increase the efficiency of WM capacity (20, 21). In addition, the episodic buffer also plays an important role in linking WM to long term memory (LTM), with retrieval occurring through conscious awareness (22). As a result, the episodic buffer would be the component of WM which makes objects consciously available. Arguably, the most important component of the Baddeley model is the central executive, which is responsible for the control and regulation of the other components. This is achieved through focusing attention, which allows specific objects to be selected in WM, or dividing attention simultaneously between multiple targets. Another executive function is the ability to switch between tasks. Finally, the central executive assists in connecting WM with LTM, both in recalling object that were previously stored in LTM, such as an individual attempting to remember a prior event, as well as selecting objects to transfer to from WM to LTM (23, 24).

The Baddeley model has been tremendously influential, however it has also been the target of criticism. For example, the function of task switching in the central executive has proven to be more complex than originally thought, with multiple stages proposed, and different modalities requiring different cognitive resources (25, 26). As a result, the prospect of a centralized task switching function appears increasingly remote (27-29). The correspondence between the model's conceptual systems and neural structures and processes is also not straightforward. One class of neuroscience models of working memory places the executive role of the model on the prefrontal cortex, whereas the subsidiary systems maintain the contents of working memory at the sensory cortices (e.g. visual, auditory) (30). However, other models dispute this division on the basis of strong evidence for sensory information being maintained within the PFC, by the same neurons that implement top-down control, thus supporting the idea of the PFC being the anatomical seat of both executive and subsidiary systems of WM (31).

In view of this criticism, and to provide an alternative framework for future studies, we propose a model of working memory inspired from the anatomical organization of the brain circuits involved in working memory maintenance and the integration of the executive

and storage WM functions into the same neuronal circuits (Fig. 1C). Areas involved in working memory maintenance are represented in color in all of the models of Fig. 1. Our model places prefrontal cortex in the center of the working memory circuit but recognizes that the sensory pathways that transmit information to prefrontal cortex are interconnected through reciprocal loops with the prefrontal cortex. In this sense, afferent cortical areas are not sufficient for the maintenance of working memory by themselves, but necessary by virtue of their connections that allow persistent activity to reverberate (and are drawn in intermediate color saturation). This proposed model supports the relative independence of different modalities, as these would activate different ensembles of neurons, though it emphasizes that this separation is not absolute within the prefrontal cortex; neurons with non-linear interaction of stimulus properties, or “mixed selectivity” have been described between domains (32). Our new model therefore represents a new starting point for the investigation of WM, building upon the models of the past.

Some recent studies of working memory, in an attempt to clarify ambiguities in vocabulary, have begun using the term “WM” to refer specifically to the process of manipulation for complex information, in contrast to “short term memory”, which is then used exclusively to denote the memory of simple stimuli (e.g. colored squares) that are maintained without any further transformation (33). Although the idea is far from unreasonable, it has created considerable confusion in the literature, as the use of “working memory” with its original meaning has persisted, in parallel.

Another unresolved conceptual debate centers on the relationship between memory and attention. At one extreme of the spectrum, WM and attention are synonymous processes: we become consciously aware of objects when we attend them, and the application of attention, such as through rehearsal, is also used to maintain objects even when they are no longer present (34-37). There is also strong evidence of overlap between the neural apparatuses of attention and WM (38). At the other extreme, attention and memory are readily dissociable, implying the possibility of attending a stimulus without maintaining it in memory (39) or storing an item in memory without even awareness of it, let alone attention (40, 41). In this review, we take the position in favor of the likely dissociation of attention and WM, even if one does not take this concept to the extreme. For example, the concept of attention is meaningful for examining stimuli that are physically present; we tend to attend specific visual stimuli even under conditions that place no memory demand. The neural correlates of attention are also evident in the processing of sensory stimuli in the early sensory cortex, despite the fact that these neurons do not maintain discharges while WM is in use (42-44). Conversely, a complete disregard for attention in WM does not seem to be possible, as attention plays a variety of roles, including the direction of focus to one of multiple stimuli held in memory (45).

It is also instructive to contrast WM with LTM. Classic neuropsychological studies suggested that patients with anterograde amnesia, most famously, patient H.M., had intact WM (46). The basis of LTM storage was thus thought to be entirely separate from WM, being mediated by long term potentiation in the hippocampus rather than persistent discharges in the PFC (47). This dichotomy has also been revisited in recent years as recent studies have recognized that the hippocampus is active during WM (48). Moreover a class

of models suggest that synaptic mechanisms make up the primary mechanism of WM, just as they do for LTM (49). Neural correlates of intermediate scales in the continuum between short-term and LTM have also been recognized (50). However, a variety of qualitative differences still remain unchallenged. Specifically, LTM does not seem to have an upper limit to its duration, lacks an upper limit to storage, and is not impaired by the addition of new items, regardless of how many were already in storage. We therefore favor their conceptual separation.

Neural Basis of Working Memory

Early neurophysiological experiments in non-human primates identified neurons that not only respond to sensory stimuli, but remain active during a period after the stimuli were no longer present; this “persistent activity” therefore provided a neural correlate of working memory (51, 52). Persistent activity has since been demonstrated in human intracranial recordings, as well (53). Visuo-spatial working memory has been a particularly fruitful model since spatial location can be varied parametrically and the activity of neurons representing each location can be studied systematically. Persistent activity in the prefrontal cortex has been shown to explain many aspects of behavioral performance in visuo-spatial working memory tasks, such that firing rate can predict whether the subject will recall the item correctly or not, or what location the subject will recall (54). However, in recent years, alternative accounts for the neural correlates of working memory have been introduced. We will group these alternatives into two categories: first, those that rely on rhythmic activity and second, those that propose information encoding without changes in mean firing rate during the delay period, that is, “activity silent” models.

Persistent Discharges

A great deal of experimental work has been centered on the representation of spatial information in working memory and theoretical research has established a concrete framework in the context of the “bump attractor model” (55-57). Models for the maintenance of object memory have been comparatively less established, considering that a near infinite number of objects can be stored in memory with no obvious network structure that can represent them in an equivalent, parametric fashion, and only a small percentage of neurons are active during maintenance of any object in memory (56). In light also of the relative segregation of spatial and non-spatial information in the brain (58, 59), we will henceforth consider these systems to be interlinked, but ultimately separate domains. In each case, we examine the evidence for persistent activity being the neural correlate of working memory and discuss arguments raised for and against it.

Spatial Working memory—Spatial working memory in both animal models and humans has been assessed through a variety of classical tasks, including the delayed response, delayed alternation, and match/nonmatch tasks. In the oculomotor delayed response task (ODR – also referred to as the memory guided saccade task) a brief visual stimulus is presented, and must be maintained in the subject’s working memory throughout a delay period, after which it is reported via an eye movement (52, 60, 61). Another common task, the delayed alternation task, also requires (hand or eye) movements to one of two

locations, alternating in successive trials and therefore requiring the preceding location to be maintained in working memory (62, 63). Individual neurons exhibit persistent activity with selectivity for different spatial locations during the performance of these tasks (Fig. 2), thus allowing the remembered location to be decoded from the activity of the neuronal population (64). The dorsolateral prefrontal cortex appears to be particularly specialized for spatial location, and recordings that sampled prefrontal neurons in a random, unbiased fashion (without isolating neurons based on their responses) found approximately 30% of neurons exhibit persistent discharges during spatial working memory tasks (65). This percentage varies between the anterior/posterior and dorsal/ventral subdivisions, with the mid-dorsal area exhibiting the highest proportion of neurons with persistent activity (66). This proportion may even be underestimated by the limited number of spatial locations typically sampled in the ODR task (e.g. eight spatial locations arranged on a ring of 10-15 degree eccentricity); experiments that used more extensive arrays of stimuli, encompassing 16 locations, reported persistent discharges from as many as 70% of all prefrontal neurons (67).

Does persistent activity represent merely motor preparation?: This question was once a common argument against persistent activity being a neural correlate of working memory (68, 69), stemming from how the location of the preceding stimulus in ODR tasks is confounded with the direction of the motor response. However, support for this argument has been critically undermined as more complex tasks have since revealed that only a minority of prefrontal neurons represent motor preparation when this factor is dissociated from stimulus properties. For example, when a task requires monkeys to make an eye movement to the location opposite to the location of the remembered visual stimulus (delayed anti-saccade task), or to a location rotated relatively to the stimulus location (rotational ODR task), the majority of prefrontal neurons that generate persistent activity represent the location of the preceding stimulus rather than the location of the impending saccade (70, 71). Moreover, persistent activity tuned for the location of a stimulus appears in the prefrontal cortex even in tasks where the stimulus does not immediately allow planning of a movement. This can be observed in the spatial delayed-match-to-sample task, where subjects are required to release a lever or press a button when a stimulus appears at a previously cued location. Prefrontal neurons generate persistent activity following the presentation of the original stimulus that is tuned for its spatial location, and not the preparation of a motor response, which cannot be planned until after the end of the delay period (72-75).

Is persistent activity merely an epiphenomenon of spatial working memory?: Strong evidence exists that persistent discharges are causally related to behavior. For example, working memory performance is significantly impaired when persistent activity is abolished via the reversible inactivation of the prefrontal cortex, e.g. through cooling (76). Under normal execution of the task, without such an intervention, trials in which persistent activity is diminished are also more likely to result in errors (52, 77). A near linear relationship between behavioral performance and persistent activity has been revealed in other tasks that parametrically modulate the difficulty of a working memory judgment (78). Lower performance of the ODR task in adolescent and aged monkeys (79) is also associated with

lower levels of persistent activity compared to young adults. Choice probability analysis, comparing the distributions of firing rates in the delay period of correct and error trials, also reveals a strong relationship between prefrontal persistent activity and the behavioral outcome of each trial (80).

Computational models provide mechanistic detail of how persistent activity influences working memory performance. Persistent activity is sustained in such models by virtue of re-entrant connections between neurons with similar tuning for stimulus properties, so that activation after afferent input is maintained in the network, which behaves as a continuous attractor (81). The bump (peak) of activity in the network determines the location recalled by the subject (Fig. 2), hence the term bump attractor (55). Drifts in neuronal activity thus account for deviations of behavior: persistent activity recorded from trials in which monkeys make eye movements deviating clockwise vs. counterclockwise relative to the true location of the stimulus yields slightly different tuning curves, implying that the peak of activity at the end of the delay period determines the recalled location (55). Similarly, the variability of a neuron's delay period activity (estimated by the Fano factor of spike counts, i.e. the variance divided by the mean) is maximal for inaccurate saccades to locations at the flanks of the neuron's tuning curve but lower for locations in the peak or tail. This finding is also explained by small deviations in saccadic endpoint corresponding to the bump of activity shifting in one direction or another, with the most rapid changes in neuronal activity occurring if the bump traverses the flank of its tuning curve rather than its peak or tail. Finally, spike-count correlations of two simultaneously recorded neurons are lowest and negative for inaccurate saccades when the cue appears between the peaks of their tuning curves. This result is also consistent with the idea that working memory inaccuracies are caused by drifts of persistent activity in the delay period; when the bump attractor randomly varies around a location between the peaks of two neurons, it inevitably causes an increase in firing rate for one neuron, but a decrease for the other. Importantly, these findings do not hold for neurons that do not exhibit persistent discharges, despite these comprising the majority of the prefrontal population (55).

Is persistent activity an “artifact of averaging”?: Some neurophysiological experiments have reported that individual neurons are only transiently representing information in the delay period (16) and persistent activity can be highly variable during the course of a trial, and from trial to trial (82). The stimulus properties can be maintained across the entire delay period only if one were to average activity across multiple trials and multiple neurons (83, 84). This led to the idea that previous reports of persistent activity were an artifact of averaging. However, these findings are entirely consistent with computational models of persistent activity. Across the population of prefrontal neurons, only a small minority would be expected to be active during maintenance of any single stimulus in memory. Variability in discharge rate during the course of the trial would be expected even among neurons that are highly active at some time point, as the activity might drift in the population. In fact, increased spiking irregularity (quantified by the coefficient of variation of the inter-spike interval) has been observed in the delay period compared to the fixation period of the ODR task (85). This otherwise puzzling finding is precisely predicted by the network models of persistent activity. Conversely, if working memory were characterized by short intermittent

bursts of high-rate firing (82), then across-trial spike-count variability (e.g., as quantified by the Fano factor) would increase dramatically during the delay periods, relative to the fixation period (86), which is inconsistent with empirical measurements (87, 88).

Are temporal dynamics of experimentally observed delay-period activity inconsistent with persistent activity?: Persistent activity is not stationary during the delay interval (72, 83, 84, 89), which does represent a contradiction of the simplest, bump attractor models, though fundamentally, models of persistent activity describe properties of a *population code*, rather than an individual neuron. Specifically, the working memory representations are encoded as a pattern of activations across a population of neurons that is not dependent upon any individual cell. Theoretical and empirical analyses have shown that stable population coding of working memory is consistent with time-varying neuronal activity (90-92). Principal Component Analysis reveals a low-dimensional representation, where stimulus location evolves dynamically in time after the cue presentation, but different locations remain constrained in separable subspaces (90).

Non-Spatial Working memory—Prefrontal neurons generate persistent discharges that represent the identity of objects held in memory, though smaller percentages of neurons are active during feature working memory than spatial working memory. Whether object working memory is localized in a different subdivision of the prefrontal cortex than spatial working memory (ventral vs. dorsal) has been the matter of debate (93, 94). At least a quantitative difference seems to be evident, with spatial information more prevalent in the dorsolateral prefrontal cortex than the ventrolateral prefrontal cortex, and this dissociation is more pronounced in the posterior rather than the anterior prefrontal cortex (95, 96).

Stimulus-selective persistent activity has been described in working memory tasks that require the maintenance of stimulus identity or features, such as shape, color, or luminance (78, 97-102). Other experiments have reported stimulus-selective persistent activity in tasks that required subjects to remember complex images, such as real objects and faces, or abstract pictures (93, 94, 103-109). Robust persistent activity has been described for the direction of motion of a random-dot stimulus that is always presented at the same location (16, 80). Persistent neuronal firing in prefrontal cortex has been observed even in the absence of performance of a task, or even learning of a task, while subjects view stimuli passively, and we are thus able to recall encountered stimuli even when we are not prompted to remember them ahead of time (54). Consistent with this finding, recordings during passive fixation reveal persistent discharges selective for faces in the ventrolateral prefrontal cortex (105).

However, the evaluation of these findings is complicated by the recent revelation that persistent activity in the prefrontal cortex also represents information beyond the characteristics of stimuli, including the abstract rules of the cognitive task subjects are required to perform (110, 111), categories (104, 112), and numerical quantities (113). Persistent activity may also represent perceptual decisions (114, 115), reward expectation (116), and sequences of events or actions (100, 117-119). Persistent firing may even represent different aspects of the same stimulus, depending on task instructions (120), and a

subset of neurons can represent multiple stimulus features and task variables simultaneously, a phenomenon known as mixed selectivity (32, 121, 122).

The realization that prefrontal activity is modulated by task factors to such extent has led to a re-evaluation of the nature of information represented in persistent activity (30). The most extreme possibility would suggest that all stimulus-selective information actually originates from task rules or categorical judgments between alternatives rather than representing the memoranda themselves. For example, in a study that required the maintenance of stimulus color in working memory, significantly more prefrontal neurons were selective to location than color, despite the fact that only color was task relevant (123). Another experiment, which required the maintenance of a sequence of stimuli, revealed a drop off in their linear classifier's ability to decode any stimuli that preceded the most recent from prefrontal activity (124).

These negative findings must be interpreted cautiously. To explain the lack of color selectivity, the activation of only a small proportion of prefrontal neurons, in the order of 5-15% (123) may be sufficient for the representation of stimulus information. It is also possible that color-selective neurons—and their persistent activity patterns—are concentrated in specific prefrontal “patches” (125) rather than be diffused across the entire prefrontal surface. Moreover, to explain the decreased decoding ability for less recent stimuli, information about multiple stimuli may be abstracted (126) so that activity representing a sequence may differ from the representation of each stimulus in the sequence, thus resulting in an apparent negative finding. The generation of stimulus-selective persistent activity in monkeys never trained to perform a task also argues against the idea that persistent prefrontal activity *only* represents tasks and rules and that stimulus information is mediated in other brain areas or through other mechanisms. Prefrontal neurons also routinely represent stimulus features even when they are irrelevant for the task at hand (78, 127, 128).

Rhythmic Models

Rhythmic activity has long been implicated in hippocampal-dependent memory, and communication between the hippocampus and prefrontal cortex, in rodents (129). In the human literature, the frequency of oscillations evident through MEG, EEG and ECoG recordings has also been associated with distinct working memory processes (130). Recent neurophysiological studies in non-human primates have begun to address more specifically what role rhythmic firing may play in working memory (131-135). The magnitude, frequency, and phase of oscillations within the prefrontal cortex and between the prefrontal cortex and other areas have been shown to be modulated depending on stimulus and task information (131, 134), thus allowing information about the stimulus held in memory or the task to be performed to be decoded based on these parameters. For example, oscillatory synchronization between LFP signals recorded from different sites within the prefrontal cortex has been shown to be modulated based on which of two task rules a monkey is performing (134). The coherence in rhythmic synchronization between neurons in prefrontal and posterior parietal cortices has also been reported to be content dependent with neurons from both areas synchronizing their firing at specific frequencies, for different stimuli held

in memory (133). The phase of rhythmic activity seems able to differentiate information representing two sequentially presented stimuli (135).

Recent studies have specifically proposed that the rate of LFP bursting in the gamma frequency range, which correlates negatively with power in the beta frequency, underlies working memory (82, 136-139). A corollary of this model is that gamma bursting pooled from error trials should be lower than that of correct trials. Unfortunately, no measures of behavior were shown to correlate with the purported neural basis of working memory during the delay interval in any of these studies. Differences in gamma bursting between correct and error trials were reported in one study (136). Critically, no differences were reported during the delay periods following the sample presentations in the working memory task used. Instead, error and correct trials were differentiated by levels of gamma bursting only during the period when test stimuli were presented and the monkeys had to judge whether they matched or not stimuli held in memory, and errors were characterized by generally higher (not lower) levels of gamma bursting (136).

Gamma bursting may still be necessary for the bottom-up input of sensory information (140). For example, in human MEG studies, gamma oscillations were demonstrated to follow visual information through the cortical hierarchy during processing into WM (141). Moreover, directing attention to sensory objects leads to sensory enhancements consistent with prior association between attention and synchronized prefrontal gamma oscillations (142-144). In any case, oscillatory activity is not incompatible with persistent activity, but instead, might reflect the underlying persistent firing and its ramifications from a distance. For example, both robust persistent activity and gamma-band rhythmicity have been reported during the delay period of the ODR task (82, 145), as well as the two-item sequential working memory task (135). Furthermore, concurrent persistent activity and gamma-band rhythmicity are observed when recordings are performed in the cortical site that corresponds to task demands: both increased persistent firing (146) and gamma-band activity (136) were captured from more ventral recording sites during an object feature working memory task. Similarly, the classic ODR spatial working memory task that generates persistent firing in dorsolateral PFC was associated with pronounced gamma bursts from the same region (82). Thus, although measures of oscillatory activity allow the researcher to sample a broader range of neuronal activity than can be performed with single or multiple unit recordings, persistent firing appears to underlie the oscillatory events captured during working memory.

Activity Silent Models

Another class of working memory models suggests that information may be maintained in working memory by neuronal population discharge patterns or synaptic mechanisms that are not evident through spiking at all. We refer to these as “dynamic encoding” and “synaptic” models, respectively, in the sections that follow.

Dynamic Encoding—Information may be represented dynamically in a neuronal population, even in the absence of rhythmicity. Specifically, the precise pattern of activation of an ensemble of neurons at each time point during a working memory task can be used

to decode the identity of the stimulus, even though overall activity during the delay period is not significantly elevated above the baseline (147). Furthermore, this pattern of neuronal activation may vary dynamically over time, with different neurons becoming active at different time points (Fig. 3). Stimuli such as those used in the Stokes et al. study are similar to those used in previous studies where persistent activity was observed (93, 107, 109). Thus, although persistent activity may have been generated, the averaging of the population might have made it difficult to detect. Dynamic activity informative about stimulus identity and task rules has been observed even when informative persistent activity is also present in the population (64, 148). Moreover, stimulus location evolves dynamically in time after the cue presentation, while different locations remain constrained in separable subspaces (90, 149, 150).

Synaptic Models of Working Memory—Activity elicited after repeated presentation of the same stimulus is typically reduced, a phenomenon termed repetition suppression (151). As a result, the level of response to a particular stimulus in the context of a working memory task, such as the delayed match to sample task, can be informative about whether it was preceded by the same stimulus or not; match suppression may signal that the sample was the same as the match (Fig. 4A-B). This suppressed response to a matching stimulus does not require persistent activity, and may be observed even when multiple seconds intervene between the sample and match (109, 152). Match suppression (or enhancement, for some neurons) is observed for stimuli matching in shape, color, and form, in spatial location, or in direction of motion, in various cortical areas, including the prefrontal, posterior parietal, and inferior temporal cortices (16, 109, 152-154). Furthermore, the extent of response difference to matching and nonmatching stimuli has predictive power over behavioral performance, as it differs systematically in correct and error trials, thus providing compelling evidence that memory performance has access to this activity (16, 155).

Computational models have been proposed that could account for such changes via mechanisms that depend on synaptic strength modifications instead of spike generation (49, 156). Such mechanisms may be mediated by calcium availability at the presynaptic terminal (Fig. 4C), whose kinetics have a time constant in the scale of seconds (49). Another possible mechanism can be seen in the processes of long term potentiation, which represents memories as changes in the synaptic architecture, rather than neuronal activity (47). Importantly, the synaptic model for working memory is not mutually exclusive with persistent activity, and cannot, by itself, account for working memory performance in other tasks such as the ODR, delayed alternation, or free recall tasks, that do not depend on a comparison of a subsequent stimulus with a prior one.

In recent years, insights about the neural mechanisms of working memory, and other cognitive functions, have been drawn from Artificial Neural Networks used to model brain processes (157, 158). Recurrent Neural Networks (RNN) in particular have been used very successfully to model complex cognitive tasks, including those related to working memory (120, 159-161). These studies have demonstrated that although it was also possible to perform simple working memory tasks by virtue of rapid changes in synaptic weights after appearance of a stimulus, mimicking the neuronal synaptic mechanisms discussed above, persistent activity also emerged spontaneously in the network, (159, 160). Importantly

however, even modestly complex tasks, such as the delayed-match-to-sample task, could not be performed by networks that did not generate persistent activity. Although not offering definitive evidence for or against the synaptic model, these findings are instructive in the limitations of synaptic mechanisms alone.

An Integrative Approach

Although we have pointed out shortcomings of activity-silent models in fully accounting for working memory, we do not wish to suggest that the underlying phenomena are not true: phenomena such as repetition suppression are robust and ubiquitous and there may be a close relation between activity-based and activity-silent mechanisms. Indeed, modeling studies that have successfully implemented these activity-silent conditions invariably require the network to be configured close to the attractor network regime (49, 162), the mechanism of persistent activity. This way, a non-specific drive can take the network to the attractor regime and reactivate a robust attractor response selected on the basis of the weak subthreshold trace. The attractor non-linearity is necessary to increase the contrast of a subthreshold signal that is fading away. Moreover, there is also the possibility of subthreshold mechanisms playing a supportive role for persistent activity in attractor networks (163). As a result, we conclude that activity-based and activity-silent mechanisms may interact synergistically instead of serving as mutual alternatives.

Recent studies examined closely the relationship of persistent spiking and activity silent mechanisms in the context of serial biases in spatial working memory (164, 165). The location subjects recall about the cue they had to remember is often biased into the direction of the cue in the previous trial (Fig. 4D), especially when the successive cues appear in close proximity (166, 167). Serial biases span the inter-trial interval between successive trials when no persistent activity survives, suggesting they are mediated by activity-silent rather than spiking mechanisms (167-170). Indeed, between successive persistent activity mnemonic codes, an activity-silent code in the PFC carried stimulus information through inter-trial periods (Fig. 4E). Increased firing rates during the onset of the following trial's fixation period revealed this latent activation as the trace reactivated: firing rate shortly after the beginning of the fixation period was tuned for the location of the previous stimulus (164). This reactivation interacted with the appearance of the following stimulus, attracting or repelling the bump generated by the appearance of the stimulus, thus creating the serial bias in memory (164, 165). This interplay could be the basis of closely associated memory storage processes operating at different time scales and serving different behavioral purposes, possibly including volitional effortful memory or occasional reactivation of recent experiences from latent traces. The overarching principle that we draw from this discussion is that synaptic mechanisms are important to the extent that they influence spike generation during working memory, which manifests itself as persistent activity.

Circuit Mechanisms of Persistent Activity Generation

The mechanisms through which persistent activity is generated and the reasons it is much more prevalent in the prefrontal cortex can be traced to a number of circuit

specializations. We will review those related to pyramidal neurons and interneurons, and to the neuromodulatory role of the monoamine neurotransmitter systems.

Pyramidal neurons and excitatory connections

The classical view of pyramidal neurons has been that they are essentially uniform across the cortex. This idea has been challenged by experimental findings demonstrating a systematic difference across the cortical hierarchy, with the prefrontal pyramidal neurons exhibiting the most extensive dendritic trees and the largest number of spines among cortical neurons (171, 172). Functional correlates of this anatomical specialization are reflected in the patterns of neuronal discharges at different areas. Prefrontal neurons receive a greater proportion of distal synaptic inputs compared to the neurons at other brain areas, with a substantial proportion of these inputs originating at distances greater than 1 mm. By contrast, the majority of inputs to posterior parietal neurons appear to originate from neurons at shorter distances, in the order of 0.2-0.5 mm (173, 174). Independent evidence for this finding is provided by anatomical studies of myelin. The MRI-based T1-weighted / T2-weighted ratio is indicative of the extent of myelin presence within gray matter and provides a measure of convergence of axonal projections (175, 176). The highest ratio is observed in the primary visual cortex and lowest (indicating most sparse connections) in prefrontal cortex (177). It has long been speculated that prefrontal neurons with similar memory fields are grouped in clusters with reciprocal connections, often visualized in anatomical tracer studies (178-181), and more extensive networks of neurons in the prefrontal cortex would explain the stability of prefrontal persistent activity.

NMDA Receptors

NMDA receptors are glutamate-gated cation channels, critical for the generation of persistent activity, as they are responsible for extending the duration of the postsynaptic depolarization through their relatively slow decay time constant (182, 183). Thus, a circuit with exclusively AMPA receptors—which produce synaptic currents with very fast decay time constants—would require unrealistically high firing rates to sustain neural activity during the delay period of a memory task (184). Experimental results further support the role of NMDA receptors in the generation of persistent activity, as the systemic administration of ketamine, a non-specific NMDA antagonist, seems to decrease the effective connectivity between prefrontal neurons, demonstrated by a decrease in the synchronous spiking between simultaneously recorded neurons (185). Persistent activity also seems to be degraded by the iontophoretic diffusion of NMDA antagonists through the cortex (186, 187), though interestingly, this result has failed to replicate in some studies (188). The area-specific expression of NMDA further underlies its role in facilitating the prevalence of persistent activity in the prefrontal cortex. For example, GluN2B—the NMDAR subunit with the slowest decay time constant, is expressed in a gradient across the primate brain, with highest levels of expression observed in the prefrontal cortex. NMDA receptor trafficking is regulated by activation of dopamine receptors (189) and dopamine innervation is concentrated in the frontal lobe (190). Thus, the NMDA receptor also represents one of the main mechanisms of dopaminergic influence on persistent activity in the PFC.

Interneuron Specialization

Prefrontal inhibitory neurons exhibit persistent activity that is thought to be essential for stimulus selectivity in working memory (60, 191, 192). Prefrontal interneurons generally exhibit higher baseline firing rates and broader tuning than pyramidal neurons (191). Their action serves to “sculpt” the spatial and temporal tuning of prefrontal neurons (193), without which stimulus-specific persistent activity is much less viable in computational models (194).

Cortical interneurons are hypothesized to form specialized networks (Fig. 5) through multiple types of GABAergic neurons for the purpose of facilitating stimulus specific persistent activity (195). Three types account for most interneurons in the cortex, those expressing Parvalbumin (PV), those expressing Vasoactive Intestinal Peptide which tends to co-localize with Calretinin (196), and those expressing Somatostatin (SST), which tends to co-localize with Calbindin. PV interneurons target the cell bodies of pyramidal neurons and when activated by their preferred stimulus, they would tend to suppress the activation of pyramidal neurons with different spatial tuning than their own and sharpen the tuning function of those with similar tuning (197). Without feedback inhibition, recurrent excitation may shift the Excitatory/Inhibitory balance to a hyper-excited state and bring the network into an unstable, hyper-excited state, which would also be deleterious for the maintenance of working memory (182).

VIP/calretinin interneurons are thought to inhibit other types of interneurons including SST/calbindin ones (198-200). Furthermore, interneuron-targeting cells are more abundant in association cortices, particularly in the prefrontal cortex, compared to the sensory cortex (201, 202). SST interneurons, on the other hand, are peridendritic-targeting cells and are thought to exhibit high spontaneous firing rates that may tonically inhibit all pyramidal neurons during the task baseline period, prior to any stimulus presentation. SST neurons would lift inhibition on the pyramidal neurons that are excited by a stimulus maintained in working memory, while other SST neurons, not recruited by the maintained stimulus, would continue to inhibit non-activated pyramidal neurons, thus suppressing both background noise, and any potential activation by subsequent, distracting stimuli (195).

Anatomical and physiological evidence supports the greater pronunciation of the disinhibiting circuit in the prefrontal cortex compared to other areas. Calretinin-positive interneurons are more numerous in the prefrontal cortex compared to visual cortical areas MT and MST (203). Moreover, interneurons with high baseline firing rate and inverted tuning (consistent with the profile of disinhibiting neurons) have also been found to be more numerous in the prefrontal cortex than in the posterior parietal cortex (204). The importance of inhibitory-to-inhibitory connections has been confirmed independently by neural-network modeling studies (161). Such connections emerge in the network as training of synaptic weights progresses and play a critical role in maintaining working memory activity. Thus, the increased presence of these circuits thus underlies the prefrontal specialization towards persistent activity.

Neuromodulatory Systems

As noted, dopamine innervation is concentrated in the frontal lobe (190) and the D1 receptor has been implicated in the generation of persistent activity. Iontophoretic application of D1 receptor antagonists, at least in large doses, compromise working memory function and erode persistent activity in the ODR task (205, 206). In contrast, D1 agonists increase activity for preferred stimuli and suppress non-preferred responses (207, 208). However the effects of dopamine receptors are complex and depend on dosage (206, 207), with differential effects on pyramidal neurons and interneurons (209). D2/D3 antagonists also suppress persistent activity, though their action primarily modulates motor responses of prefrontal neurons (210). Computational and experimental studies suggest that the overall effect of dopamine is to improve the signal-to-noise ratio of persistent activity (211-214).

Other monoamine neurotransmitter systems, although not exclusive to the frontal lobe, influence persistent activity. For example, blockade of α_2 adrenergic receptors erodes spatially tuned persistent activity and α_2 agonists enhance prefrontal persistent activity (215). Similarly, cholinergic agents (both muscarinic and nicotinic) are known to influence persistent activity. Muscarinic antagonists impair working memory performance (216) and suppress prefrontal persistent activity (217). Nicotinic α_7 -nACh receptor agonists enhance and antagonists inhibit prefrontal persistent activity (218). Capitalizing on the known effects of the cholinergic system some recent studies have relied on endogenous activation of the cholinergic system through electrical stimulation of the Nucleus Basalis of Meynert to improve working memory (219, 220).

Localization of Working Memory Activity in the Brain

Persistent discharges are not an exclusive property of the prefrontal cortex. Neurons in premotor, parietal, cingulate, and temporal association areas also generate robust persistent activity, as do subcortical structures including the basal ganglia and the medio-dorsal nucleus of the thalamus (221, 222). Models of working memory relying on activity-silent mechanisms have suggested that working memory is localized in sensory areas, as early as the primary visual cortex (223). We will review the evidence for working memory representation in all of these brain areas.

Posterior Parietal (PPC) and Inferior Temporal (IT) cortex

The posterior parietal and inferior temporal cortex represent two main cortical afferents to the prefrontal cortex (221). Neurons in posterior parietal cortex share many functional properties with neurons in the dorsolateral aspect of the prefrontal cortex, to which they are interconnected (224) and both regions are activated simultaneously in human imaging studies of spatial working memory (225-231). Moreover, neurons in the posterior parietal cortex also generate persistent activity through a virtually identical percentage of neurons compared to the dorsolateral prefrontal areas during the ODR task (232). The remembered locations of visual stimuli can be decoded from the persistent activity from either area, independent of a planned motor response (233, 234), thus implying that spatial working memory representations may be either redundant or distributed across multiple areas at once.

Similarly, the IT cortex shares a number of physiological properties with the ventrolateral prefrontal cortex and both exhibit persistent activity that encodes the features of remembered stimuli during the delay period of object WM tasks (235-241). As noted in previous sections, the simultaneous activation of the prefrontal cortex and its interconnected areas during working memory tasks has led to the implication that the prefrontal cortex may not necessarily comprise the complete working memory representations, but would only hold the focus of attention, or other top-down signals, while the actual contents of working memory would be represented in the posterior parietal and inferior temporal cortex (242-246).

One line of evidence suggesting that the prefrontal cortex maintains stimulus information not present in its afferent areas comes from memory tasks that require maintenance in memory of an original item through sequential presentation of distracting stimuli, such as the object and spatial versions of the delayed match to sample task. In object delayed-match-to-sample task, persistent discharges of IT neurons are interrupted by non-matching, distractor stimuli presented after the sample (237). Similarly, persistent posterior parietal discharges represent the most recent stimulus location and are disrupted by distracting stimuli in spatial delayed-match-to-sample tasks (233). In contrast, responses in the prefrontal cortex are able to represent the actively remembered object (109) or spatial stimuli (72, 247, 248).

More recent studies have somewhat qualified these findings, demonstrating that differences between IT/PPC and prefrontal neurons are quantitative rather than qualitative (72, 154), and that in some tasks, prefrontal neurons may display greater responses to distractors than actively remembered stimuli (249, 250). Nonetheless, performing the task reviewed in the previous paragraph seems impossible based on the activation of the posterior parietal or inferior temporal cortex alone. Accumulating studies support the possibility of the prefrontal and PPC/IT cortex simply being specialized for different aspects of working memory, among other cognitive functions (249-254).

The primacy of the prefrontal cortex in working memory behavior is perhaps most vividly demonstrated in inactivation studies. Cooling experiments, which reversibly inactivate the underlying cortex by lowering its temperature, demonstrate much greater decreases in ODR performance after prefrontal cooling compared to posterior parietal cooling (255), even when the inactivated areas have similar delay period activity (232). The results of these studies parallel the effects of reversible inactivation of the frontal eye fields via muscimol injections, which produce a significant impairment in memory-guided saccade performance (256, 257). In contrast, virtually no impairment was observed after the muscimol-induced inactivation of the posterior parietal cortex (255, 258, 259), even though posterior parietal inactivation produces consistent deficits in tasks that require attention or selection between multiple stimuli (258, 260-262). Moreover, small lesions to the dorsolateral prefrontal cortex also produce impairment in working memory performance for remembered stimuli in the contralateral space, an effect termed a “mnemonic scotoma” (263, 264).

Visual Cortex

Recent human imaging studies have applied Multi-Variate Pattern Analysis (MVPA), examining the simultaneous pattern of activation of multiple voxels to different task conditions (265), in order to successfully decode working memory content from the primary (223, 266, 267) and extrastriate visual cortex (268-270). MVPA analysis could not decode information from the prefrontal cortex, nor fully account for behavioral performance in the task (223, 268). As a result, some imaging studies have even claimed that the size of the primary visual cortex alone may be the best predictor of working memory ability (271).

However, there are critiques that may call these results into question. A tacit assumption when comparing the results of MVPA analysis is that the structure of the voxel (typically in the order of 3mm x 3mm x 3mm) would be equivalent across different cortical areas. This may not be the case. Unlike the well-documented topography of visual space in the primary visual cortex, no retinotopic map (or other overarching organizational principle) has yet been revealed in the prefrontal cortex, and the same retinal position is represented multiple times across the prefrontal surface. Precise stimulus location information therefore seems to be represented on an extremely fine spatial scale, with the entire visual hemifield possibly represented in prefrontal modules with a surface no larger than 0.5 x 0.5 mm (61). Voxels averaging cortical volumes an order of magnitude larger are thus likely to eliminate stimulus information and fail to decode the information held in working memory, even if this is robustly represented in activity of individual prefrontal neurons. Positive fMRI results, decoding features of a remembered stimulus directly from prefrontal cortex have also been published: the orientation of a grating was successfully decoded from the prefrontal cortex (272). These results thus argue directly against models of working memory that postulate a solely top-down control role for the prefrontal cortex, and place feature storage networks in the visual cortex.

MVPA methods still yield undeniably positive findings of fMRI imaging hence it is important to consider neurophysiological activity in the visual cortex during working memory. By some accounts, most if not all areas of the visual cortex produce neurophysiological activity during working memory (273). A more careful view of these results however, leads to a more nuanced view. Some studies have reported stimulus-selective activity in a small percentage of V1 neurons during working memory (274). However, activity that outlived the visual stimulus tended to be characterized by a suppression of firing rate below the baseline, rather than elevated persistent activity. Importantly, the paradigm used in these experiments relied on the presence of a background grating that remained visible after the foreground stimulus was no longer present, providing physical stimulation onto V1 neurons. Stimulus-selective activity in V1 during working memory is likely due to top-down projections from higher associative cortices, since V1 activation appears first in superficial layers (275). The relatively “quiet” background activity in the V1 allows the observation of this subtle backwash from higher cortical areas, while the higher cortical areas themselves may be too noisy to detect these small signals. In addition to this backwash, fMRI activation may also be detecting pre-synaptic activation of V1 neurons from higher cortical areas (276). It is also unclear whether V1 activity can be predictive of behavior in working memory tasks as this modulation was present for

both correct and incorrect trials (274). On the other hand, studies comparing activity in three cortical areas in the same animals required to remember the direction of motion of a random-dot display, found virtually no persistent discharges in visual area MT, but robust activation in parietal area MST, in addition to prefrontal persistent activation (80). This suggests an abrupt generation of feature-selective persistent activity in areas beyond the visual cortex. A more recent review of cortical areas generating persistent activity, including both positive and negative findings, revealed activity concentrated in the prefrontal cortex and areas directly connected to it (44).

Subcortical Structures

Basal ganglia—In primates, the basal ganglia are formed by the striatum, the globus pallidus (Gpe, Gpi), the substantia nigra (SNc, SNr) and the subthalamic nucleus. Cortico-striato-thalamo-cortical loops, which involve various subdivisions of the PFC, are then used to project thalamic and Gpi/SNr outputs back to cortical areas (277-280). Several nodes of these loops are targets of dopaminergic cells. The presence of these inputs is therefore implicated as a possible mechanism through which these loops may contribute to working memory related behaviour beyond purely motor control, binding sensory, motor, and motivational cortical information (281). Substantial persistent activity has been described in the basal ganglia (282-286). As in prefrontal cortex, visual and persistent activity during memory delays in the caudate nucleus reflects combined visual and motivational (reward expectation) information needed to perform working memory tasks (284). The basal ganglia are thus implicated in working memory as a possible support system for prefrontal-dependant maintenance.

Thalamus—The thalamus is a key structure in both cortico-striato-thalamo-cortical loops, as well as direct cortico-thalamic systems. These latter systems may combine with cortico-cortical connections to implement collective computation and large-scale fast communication, which may ultimately be the mechanism of cognitive binding (287-289). Several thalamic nuclei connect with associative cortices, and most importantly, the mediodorsal nucleus (MD) and the medial nucleus of the pulvinar (290-294). Neurophysiological studies in MD have shown typical persistent activity during memory tasks (295-297), though more neurons are activated by the motor aspects of the task (297, 298).

Hippocampus—The hippocampus is well accepted to play a critical role in the consolidation of working memory into long-term memory. This role implies tight communication with the prefrontal cortex. Indeed, hippocampal neurons are readily activated during working memory tasks, such as the delayed-match-to-sample task, and exhibit stimulus-selective persistent discharges (299).

Long-range loops

The review of areas where persistent activity has been observed hints at some general principles. The prefrontal cortex and the brain areas to which it is directly connected to forms a series of parallel, long range loops comprising the association cortex of the parietal and temporal lobe, and cortico-striatal, and cortico-thalamic circuits, where persistent

activity is readily detectible. Recent theoretical studies suggest that inter-areal reverberation is essential for the emergence of persistent activity, even when some of the component areas are incapable of generating persistent activity on their own (300). In this context, the prefrontal cortex can be viewed as a core component but not as the exclusive site of working memory maintenance.

Functional Organization of Working Memory Within the Prefrontal Cortex

Anatomical organization

The prefrontal cortex can be divided into a lateral, a medial and an orbital aspect, the former of which aspect can be further distinguished into a dorsolateral and a ventrolateral subdivision (221). Multiple cytoarchitectonic areas make up the lateral aspect of the macaque prefrontal cortex, including areas 8a (encompassing the Frontal Eye Field) and 45, which lines the superior and inferior banks of the arcuate sulcus respectively, area 8b, which lies just medial to the arcuate, areas 9 and 12 in the superior and inferior convexities of the cortex respectively, area 46, which lines the banks of the principal sulcus and finally, area 10, which covers the frontal pole (301). Area 46 has since been further subdivided along its mediolateral axis, into areas 46dr, 46d, 46v and 46vr, which line the medial rim, the medial and the lateral banks of the principal sulcus and the lateral rim of the principal sulcus, respectively (302), and along its anterior-posterior axis into a caudal aspect—sometimes referred to as area 9/46—and a rostral one(303). fMRI studies have implied the possibility of additional areas based on functional connectivity at rest (304).

Anatomical studies have found that projections from the posterior parietal cortex terminate primarily in the dorsal prefrontal cortex (areas 8 and 46, including both banks of the principal sulcus), while the inferior temporal cortex projects to areas 12 and 45 of the ventral prefrontal cortex, thus suggesting a relative segregation between the areas. (305-307). A relative segregation of inputs has also been observed across the anterior-posterior axis, with areas higher in the sensory and limbic hierarchies projecting to more anterior prefrontal subdivisions, while lower areas likewise project to the more posterior subdivisions instead (308-310).

Localization of stimulus selectivity

Guided by the pattern of anatomical connections, early neurophysiological studies suggested that spatial and object working memory were subserved by the dorsolateral and ventrolateral prefrontal cortex, respectively—a “domain-specific” organization that was thought to extend from the dorsal and ventral visual streams (59, 105, 106, 311, 312). However, opposition to this proposal has noted that after a monkey is trained in tasks that require stimuli location and identity, neurons with selectivity to these factors can be found in all areas of the prefrontal cortex. (93, 313). This would imply that the apparent functional specialization observed in the earlier studies was actually the result of task requirements, thus leading to the proposal of a new “integrative” model, where task requirements and cognitive process engaged are the primary principle of organization rather than the type of stimulus representation (93, 230, 313).

Spatial selectivity

More recent studies have reexamined prefrontal selectivity for different stimulus properties by comparing dorsal and ventral prefrontal responses to the same stimuli. Spatial selectivity (Fig. 6). proved a strong predictor of whether a neuron was recorded in the dorsal or the ventral prefrontal cortex, with dorsolateral prefrontal cortex neurons displaying significantly higher spatial selectivity, regardless of whether or not the monkeys were trained, passively viewing the task, or actively maintaining the stimuli in their working memory (56, 95, 314). Ventrolateral prefrontal neurons also displayed substantial spatial selectivity with well localized receptive fields, but only to a comparatively lesser degree (95, 314). Importantly, a gradient of stimulus selectivity is present along the anterior-posterior axis, with the posterior areas containing the most highly selective neurons; anterior neurons exhibit little selectivity to stimuli per se, but are more likely to represent task variables (314). Spatial selectivity also exhibits a strong temporal component. The proportion of spatially selective neurons, for example, peaks early after the appearance of the stimulus and diminishes with time; response latency also differentiates the highly selective posterior areas from the less selective anterior ones (314, 315). A functional specialization between (posterior) dorsal and ventral areas is also most evident for peripheral stimuli. Experiments that positioned stimuli 10-14 degrees of visual angle away from the fovea were able to detect the dorsal and ventral gradient of specialization, while studies that tested stimuli within 4-6° degrees from the fovea found little or no differentiation between dorsal and ventral prefrontal activity (93, 315).

Studies of artificial manipulation of neuronal activity provide a second line of evidence regarding the functional specialization of the prefrontal cortex. For example, temporary inactivation experiments e.g. injections of the GABA_A agonist muscimol, or lidocaine, or cooling of the underlying cortex in the dorsal prefrontal areas, including the Frontal Eye Field, causes spatial working memory deficits that localize in the contralateral visual field, (76, 248, 256, 257, 263, 316, 317). In contrast, inactivation of the ventral prefrontal cortex does not impair spatial working memory (318).

How visual space is organized in the prefrontal cortex remains unresolved. No obvious, retinotopic organization of visual space has been described across the surface of the prefrontal cortex, or with dense electrode areas placed in posterior dorsal areas with strong overall spatial selectivity (Fig. 7A). Some indications exist for at least some local organization, however (Fig. 7B). Electrode penetrations into the principal sulcus suggest an orderly progression of receptive field locations (319). Some organization has been revealed across cortical layers, at least in terms of time course of activation (Fig. 7C): neurons activated by visual stimuli are more likely to be encountered in intermediate layers; neurons with delay period activity in superficial; neurons with motor responses in deep, though this relative segregation is quantitative rather than qualitative (69).

Feature selectivity

In monkeys trained to perform behavioral tasks, dorsolateral prefrontal neurons (including the frontal eye fields) exhibit substantial selectivity for object shape, no less than that of ventrolateral neurons (95, 320, 321), and mirroring the selectivity observed in the

posterior parietal cortex (322, 323). Interestingly however, if the monkeys are not trained, shape selectivity is higher in the ventrolateral prefrontal cortex instead (95). Unlike spatial location, which can be studied systematically with parametric variation, the virtually infinite variety of shapes prevents the detection of a clear-cut dichotomy in shape selectivity between the dorsal and ventral prefrontal cortex. However, neurons with high feature selectivity will only respond vigorously to a limited set of stimuli and are likely to produce uniformly weak responses to stimuli drawn from a small set, failing to differentiate between them. This precisely matches the response patterns of inferior temporal neurons (324-327) and ventrolateral neurons. By contrast, when neurons were tested with stimulus sets requiring very narrow shape selectivity, such as faces and complex objects, neurons distinguishing between such stimuli were localized exclusively in the ventral prefrontal cortex (105, 106).

Selectivity to stimulus color follows a similar pattern. Only a small proportion of the total prefrontal populations exhibit color selectivity, in the order of 5-15%, similar to the proportional selectivity in posterior parietal neurons (123, 328). Among neurons that respond to colored squares, selectivity for different color is not significantly different in dorsal than ventral prefrontal cortex (314). Combined fMRI studies and neurophysiological studies in the temporal lobe have suggested that neurons selective for faces, other shapes, and colors are clustered at distinct patches of cortex (329-331). A handful of studies have suggested a similar clustering for color-selective neurons in the prefrontal cortex as well, thus strongly implying the precise anatomical localization of color representations in working memory (125).

Plasticity of stimulus representations

Plasticity of neural responses also exhibits differential localization in the prefrontal cortex (Fig. 6). While training increases the number of responsive neurons in the ventral and anterior prefrontal cortex, the posterior dorsal prefrontal cortex exhibits the weakest effects of plasticity, with robust selectivity for spatial stimuli being present both before and after training (95, 314). The emergence of robust selectivity after training would thus seem to indicate that anterior and ventral responses are more sensitive to task variables and cognitive factors—particularly reward conditions—rather than stimulus properties, and this capacity for plasticity has been linked to the action of dopamine D1R receptors (332). The great plasticity of the prefrontal cortex cannot entirely erase the pre-existing functional specialization gradient for spatial and object information between dorsal and ventral prefrontal cortex, in their posterior aspects. Moreover, training does not exert an equal effect on all areas of the prefrontal cortex, thus implying that the effects of training on any given neuron would depend upon that neuron's anatomical location.

Functional implications of dorsal and ventral prefrontal inactivation

Although selectivity to different types of stimuli is revealing, the ultimate functional role of an area can be probed by examining the consequences of its activity on performance of working memory tasks. Thus, experiments manipulating neuronal activity can be very informative on the functional roles of prefrontal subdivisions. Temporary inactivation experiments e.g. injections of the GABA_A agonist muscimol, or lidocaine, or by cooling of

the underlying cortex, during working memory are consistent with the specialized stimulus selectivity of dorsal and ventral PFC neurons. Inactivation of dorsal prefrontal areas, including the Frontal Eye Field, decreases performance during spatial working memory tasks (76, 248, 256, 257, 316). Similarly, limited injections of muscimol in the dorsal prefrontal cortex produce spatial working memory deficits that localize in the contralateral visual field. This “mnemonic scotoma” effect is also produced by small, focal lesions (263, 317).

Inactivation of a dorsal area, the Frontal Eye Field, has negligible effects on object working memory (333), even when monkeys are tested with objects for which Frontal Eye Field neurons were shown to exhibit broad but significant selectivity in the same object-working memory task (320). In contrast, inactivation of the ventral prefrontal cortex impairs the ability to locate objects based on remembered features, but not on spatial location (318). Location along the anterior-posterior axis is also critical for the effects of lesions. It was lesions in the anterior aspect of the ventral prefrontal cortex (area 47/12) that failed to produce deficits of feature working memory (334), and in the posterior that succeeded (318).

Although our review focuses on spatial and object working memory, we do not wish to suggest that these are the only functions of the dorsal and ventral prefrontal cortex, respectively. The representation of task rules and associations reviewed in the neurophysiological studies of the previous section fares prominently on the functional consequences of lesion studies. In this case too, the effects of lesions are dissociable between dorsal and ventral prefrontal cortex. Thus, lesions in the posterior-dorsal PFC impair performing tasks relying on associations (e.g. green light means press left button, red light means press right), even after the basic rule of the task has been acquired (335). Lesions of the mid-dorsal PFC result in deficits in tasks such as the delayed non-match to sample, involving presentation of a stimulus and after a delay period, the same stimulus plus a new one, requiring selection of the newly added item, particularly if this is not entirely novel object but it is one the monkey is familiar with (335). Damage to the ventral prefrontal cortex does not produce impairments of performing tasks requiring recognition or simple recall; its effects become apparent in tasks requiring active retrieval of information from memory, such as free-recall tasks (336). The ventral prefrontal cortex also appears to be essential for learning a task by trial and error, and reversal learning, which requires learning new associations within a session (334, 337, 338).

Neural Correlates of Working Memory of other Modalities

Although visual working memory is arguably the most popular modality in current research, the examination of additional sensory modalities is also highly instructive. In particular, the comparison between modalities reveals common themes on the neural basis of working memory and informs the debates on the mechanisms of its generation and its localization in the brain. We will examine primarily somatosensory and auditory working memory, since these are the two modalities most studied in non-human primates; reports of olfactory and gustatory working memory will be briefly reviewed in the last part of this section.

Somatosensory Working Memory

Somatosensory information is processed by the primary somatosensory cortex (SI), which is situated in the postcentral gyrus. This information is then relayed to the secondary somatosensory cortex (SII) and higher order somatosensory areas including areas 5 and 7b in the parietal lobe and the insular cortex (Ig), before ultimately being projected to the frontal lobe. The ventral rim of prefrontal area 46 is anatomically connected with parietal lobe regions, with the strongest connections stemming from area 7b (306). Two paradigms have been popular in the study of somatosensory working memory. In one series of studies, subjects were required to remember the frequency of a vibratory stimulus and to determine if a second vibratory stimulus was of higher or lower frequency (as in Fig. 8, bottom panel). Neurons in the prefrontal cortex and other frontal areas (e.g. premotor and pre-supplementary motor area) have been shown to remain active in the delay period after the presentation of the first vibratory stimulus, and furthermore, their firing rate is graded, depending on the frequency of the stimulus (339). Similar to what has been observed in visual working memory neurons, persistent somatosensory activity typically exhibit dynamics, with activity ramping up or down towards the end of the delay period, but vibratory stimuli are still represented in a stable subspace (90, 340-342). Frequency-selective persistent activity has also been described for cross-modal tasks (as in Fig. 8), where the subject has to maintain in working memory the frequency of a tactile stimulus and compare it with that of an auditory one (343, 344).

In a second popular paradigm to study somatosensory working memory, subjects are presented with a rod, out of view, whose surface is an embossed grating or other patterned texture, and the subjects have to maintain this tactile information in working memory in order to determine whether a subsequent stimulus presented in the same fashion is the same or different. Neurons in the PFC remain active after the presentation of such stimuli, and individual neurons exhibit selectivity for different patterns (340). In this case too, subjects can perform cross-modal judgements, e.g. comparing the orientation of the tactile stimulus with that of a visual one (340).

These experiments reinforce some of the ideas discussed in the context of the role of persistent activity in visual working memory. Individual neurons readily generate persistent activity, which is specific for the properties of the stimuli maintained in memory and not related to motor preparation (which cannot be planned at the time of the first stimulus presentation) or their spatial location (which never varies in the somatosensory paradigms mentioned), so as to instantiate top-down control.

The localization of somatosensory working memory has also been debated. Beyond the frontal lobe, haptic-selective persistent activity has also been demonstrated unequivocally in brain areas directly connected with the prefrontal cortex, most importantly the posterior parietal cortex (345). In sensory cortical areas, results have been mixed. Tested with the vibratory working memory paradigm, neurons in SII have been shown to discharge for a brief period after the offset of the stimulus, though these responses quickly decay (346). Frequency-selective vibratory information was not found in SI in these experiments (347). On the other hand, a small population of SI neurons exhibit sustained haptic-selective activity over the entire delay period for textures of remembered objects (348). We may

therefore conclude that, in analogy to the visual cortex, primate somatosensory areas are not entirely uncoupled to working memory. Persistent activity driven by somatosensory stimuli is more prevalent and more robust in the prefrontal cortex, though traces of persistent activity can be found under some conditions as early as in the primary somatosensory cortex. It appears therefore that a gradient of importance exists, with frontal areas simply being the most crucial for stimulus maintenance, being active across conditions, and enabling judgements such as cross-modal comparisons of remembered stimuli.

Auditory Working Memory

The auditory cortex is organized in a central primary auditory cortex core (A1) that lines the ventral bank of the Sylvian fissure and is surrounded by an area of higher-order auditory cortex, known as the belt cortex (A2). A tertiary auditory area, the parabelt cortex (A3), is located lateral to this belt. In analogy to the segregation of object and spatial information in the visual pathway that extends into the PFC, the rostral and caudal belt have been identified as processing pathways for the identity and spatial location of auditory stimuli, respectively, and their projections have been shown to terminate in discrete subdivisions of the frontal lobe (349-352). Specifically, the rostral belt projects to the inferior convexity (areas 12 and 45), a region that also receives visual object information from the inferior temporal cortex, whereas the caudal belt cortex targets the caudal aspect of area 46 and area 8, a region that also receives visuo-spatial information from the posterior parietal cortex (353-355).

Neurons in the prefrontal cortex generate persistent discharges in auditory working memory tasks that represent remembered sound features, such as the identity of a monkey vocalization, or the spatial location of the sound source (356-360). Auditory information that is maintained in memory and generates persistent activity allows cross-modal judgments as well (359). Persistent activity representing the identity of multi-sensory, audio-visual stimuli has also been described (361). In analogy to visual WM, persistent activity in auditory WM tasks also encodes more abstract information e.g. categories of sound stimuli (362). Activity-silent mechanisms of auditory WM have also been identified in the PFC, most importantly differential responses to the same stimulus when it appeared as a match or a nonmatch in WM task (363).

The localization of auditory memory in the brain has also received substantial research attention. Beyond the prefrontal cortex, persistent activity for auditory stimuli has been described in the superior temporal cortex (364). The primary auditory cortex also seems to possess a small population with stimulus-selective persistent activity (365, 366) (367), though the result has not been confirmed across all studies, and in some auditory working memory paradigms, persistent activity was not detected in the primary cortex (368). Strong evidence exists for a causal role of the prefrontal cortex in the maintenance of working memory; reversible inactivation experiments have revealed profound impairments in auditory and audio-visual working memory tasks (369).

Gustatory and olfactory pathways

Gustatory cortical areas include the primary gustatory cortex in the postcentral gyrus, and secondary areas such as the insula and precentral opercular areas. These areas largely

project to areas 12 and 13 of the orbitofrontal cortex (370). The same orbital frontal areas are also targeted by the olfactory (pyriform) cortex, while the entorhinal cortex receives input directly from the olfactory bulb (371). Although few neurophysiological studies have employed gustatory and olfactory cues for tasks that engage working memory in primates, there have been reports of persistent responses for gustatory stimuli, primarily in the secondary gustatory cortex and in the orbitofrontal cortex (372, 373). In one set of experiments, neurons in these areas sustained firing while the subject had to remember the identity of the gustatory stimulus (water or salt) in order to plan a response. Such sustained responses were absent from the primary gustatory cortex, the neurons of which encoded gustatory information during the stimulus presentation (372). In another set of experiments, neurons in the orbitofrontal and gustatory cortex generated persistent discharges representing the identity of gustatory stimuli (different types of fruit juices) in the context of a match/nonmatch task. Some neurons with persistent gustatory discharges were also encountered in the dorsolateral and ventrolateral prefrontal cortex, though these were less frequent (373).

Working Memory Capacity

Limited capacity is a defining feature of working memory and is widely viewed as a predictor of general cognitive ability (13, 38, 374-376). The limits of capacity are not exclusive to working memory storage alone, but also include the amount of perceptual information that we can process at any given time, traditionally exemplified in change blindness demonstrations (377). Further studies have suggested that this apparently blindness may be due to unattended objects being stored as ensemble statistics, where multiple, single item measurements are combined to form a singular group (378). This would imply that only attended objects are directly stored in WM capacity, while all others would be stored as single or multiple ensemble statistic sets (379). We examine behavioral correlates of working memory capacity first, and conceptual models that have been introduced to account for them, before we delve into the neural correlates that underlie the storage of information in WM and its limitations.

Behavioral Correlates of WM Capacity

Experimental tasks devised to quantify how many items a subject can maintain in memory often rely on “change detection”: subjects are typically presented with a set number of items and after a delay period they are asked to report if one of the items presented for a second time was identical to the first (single probe test), or if a whole display presented for a second time is identical to the first. In a set of experiments that used a change detection task to present arrays of colored squares, oriented lines, variously sized lines and other similar stimuli, participants were generally found to maintain up to three distinct items without difficulty or error (20). However, any arrays that demanded capacity for additional items beyond this point seemed to cause a systematic decrease in performance, demonstrating the extreme limitation of human working memory capacity.

Analytical models of capacity have been developed that can provide estimates of the items a subject can maintain in memory correctly. Two popular examples include the Pashler and

Cowan formulas for WM capacity, which both assume that WM is carried out through a limited number of discrete units in the change detection paradigm (20, 380):

$$\text{Pashler: } k = N \left(\frac{h - f}{1 - f} \right)$$

and

$$\text{Cowan: } k = N(h - f)$$

In both formulas, variable “h” represents the number of hits in the change detection paradigm, variable “f” represents the false alarm rate, variable “N” represents the set size of remembered material that is demanded by the paradigm, while variable “k” represents the estimated measure of capacity. Cowan’s formula applies primarily to single-probe experiments of change detection; Pashler’s formula is most applicable in whole-display paradigms (381). A caveat is that these models predict perfect performance for any set size that is smaller than capacity—an assumption that it is almost always incorrect due to the fact that occasional errors may occur at any set size due to lapses, e.g. correctly recalling but mis-clicking on the task computer, or circumstantial distractions of external stimuli or internal thoughts.

The definition of the units of capacity is also tricky, as it is well known that WM may actually be carried out as the integration of multiple items into “chunks” (e.g. groups of digits). Alternatively, it is also possible that all features may be stored as a continuous ensemble, rather than being “chunked” into discrete units (further discussed below). A “unit” may also include multiple features of the same object: If the task requires storage of color and orientation simultaneously, in a conjunction task, performance does not decrease over maintenance of one feature at a time (375). On the other hand, spatial locations of objects seem to be stored more reliably and automatically in WM, even if they are not task relevant (382). The estimated capacity may also dependent on the complexity of the objects stored in memory and the relative similarity of items in the display (21, 383).

Discrete Slots vs. Continuous Resource

Whether WM capacity is bound as a set number of discrete units—typically referred to as “slots”—or instead, if capacity is a flexible resource that allocates items continuously, and therefore, without a limit fixed to a single number is a matter of debate (375, 384). The discrete units model can be traced to early models of WM capacity, introduced by Miller in 1956 as the “magic number seven” for the number of objects that could be held in WM at any given time, although more recent estimates suggest an even lower number of four or so objects (20). Evidence in favor of the former model comes from color recall tasks in which subjects indicate the closest shade on a spectrum wheel. The error distributions are consistent what would be expected from discrete units, with each additional item in WM load decreasing the probability density of correct responses until capacity is reached, and performance completely collapses (13, 385). Moreover, beyond the behavioral evidence,

EEG recordings also revealed a steady increase in mean amplitude for each item that was added to capacity, which would eventually plateau when capacity was reached (385).

Flexible resource models point out the inverse relationship between WM capacity and object complexity. An extreme view of the discrete units model would predict that the number of items that can be stored in capacity would remain constant regardless of the complexity of the stored objects. However, this has not been experimentally supported. The addition of new features or objects increases the net complexity and decreases apparent WM capacity accordingly (384, 386). Results of the error distributions in color recall tasks have also been revisited. The Target Confusability Competition model suggested that errors are primarily driven by the similarities between alternatives (383). Behavioral evidence from a color detection task indicated that stored objects with greater similarity to the target seem to receive a boost in their likelihood to be selected, thus providing a possible explanation for the abrupt collapse in performance that occurs when capacity is reached (383).

For both discrete units and flexible resource models, there seems to be a link between attention and WM capacity, as prior studies have suggested that individuals who possess a higher WM capacity may be able to control and maintain the intensity of their attention on a specific object more effectively than individuals with low WM capacity (374, 376). However, despite the general rule of WM capacity decreasing as remembered objects become more complex, not all types of WM seem to be affected equally in this manner. A critical example can be observed in a recent study that applied a recall task of increasing load to examine the effects of increasing task complexity on WM capacity, ultimately finding that verbal WM was not affected (387). As a result, verbal WM would seem to be more aligned with the expected results of the discrete unit models of WM capacity, despite these models being opposed by other studies, such as the color change detection task of Bays et al (386).

Electrophysiological Correlates of WM Capacity

The number of items stored in working memory manifest themselves in electroencephalographic (EEG) activity that persists throughout the entire period in which the content is maintained in WM (36, 388-390). The Contralateral Delay Activity (CDA), a negative-voltage wave over posterior contralateral electrodes increases monotonically rising in amplitude as the more items are added (Fig. 9), eventually flattening into an asymptote when the subject individuals reached their respective WM capacity limits (390). In fMRI studies, the posterior parietal cortex, particularly in the intraparietal sulcus, seems to be particularly important, as the levels of observed BOLD activation in this area best tracks WM capacity (33). Interpretation of these macroscopic measures of brain activity hinges on the fundamental mechanisms of working memory maintenance. Increased content may simply be more susceptible to drifts in activity, without reducing the activity of the neural ensembles that represent each given object (391). Alternative models of WM capacity rely on the number of EEG cycles that can be simultaneously maintained, with each cycle representing a different object, as each cycle would represent the firing of a different neural ensemble (392, 393).

Non-human primates are capable of mastering multiple-item working memory tasks (394-397) and it has been possible to investigate the representation of multiple-stimulus information in memory by single neurons (146, 395, 396). In the context of the bump attractor model (Fig. 10), the activity representing each of multiple stored items held in memory can be thought of as a separate bump (398). As more items are added to memory, the total population activity would be expected to increase. However, when the capacity of the system is exceeded, the persistent activity representing some of the stimuli decays and the item cannot be recalled at the end of the delay period. In practice, experimental studies have not produced such a clear picture of the neural correlates of capacity. Prefrontal neurons have large receptive fields, and firing rate of individual neurons quickly saturates when more than a single item is maintained in memory, as does the information that can be decoded from their activity (396). Across the population of neurons, activity in the delay period does increase as additional items are stored in memory (Fig. 10), by virtue of more neurons being recruited (399). However responses to multiple item displays is not the linear summation of responses to individual items, which complicates the decoding of items held in WM (123).

Working Memory Duration

As with capacity, WM is defined, in part, by its limited duration, which is commonly believed to span along a range of seconds. Certainly, there is a significant separation between this, and the hours-to-decades long storage duration that has been observed to occur in LTM (400-403). However, there is a great deal of debate in regard to where one form of memory begins and the other ends. We will therefore attempt to reinitiate this line of inquiry by examining the neural correlates of WM duration across the visual, auditory, and tactile modalities, the effects of rehearsal, and the underlying nature of WM duration in regard to whether this property is limited by the decay or interference of stored information.

Distinguishing Maintained WM Duration from LTM

Due to the ambiguity in separating WM duration from LTM, many researchers have suggested that at least certain aspects of WM duration may actually be activated LTM (404-406). This hypothesis is supported by the wide degree of overlap between the brain areas that are involved with LTM compared to the brain areas that are involved WM, particularly in the prefrontal cortex, and in the case of auditory WM, the hippocampus (405, 407). Aging also affects these areas simultaneously, resulting in similar deficiencies in both LTM and WM (408). Moreover, certain models of WM, such as Baddeley's phonological loop, would require the representation of information in LTM, regardless of whether or not WM is a separate and distinct function (1, 409). However, the activated LTM explanation for WM duration also faces a degree of challenge, particularly in the rapid loss of information that is often observed in WM, compared to the extremely long lasting information in LTM (402). If WM were truly activated LTM, we would expect that subjects would be able to retrieve information previously discarded from WM, even a significant period of time later. We would also expect that the impairment of LTM would result in the impairment of WM, though this was also refuted by prior examinations on the effects of head trauma (401).

Furthermore, despite being involved with similar brain areas, LTM and WM seem to rely on very different molecular mechanisms (400, 410). LTM, for example, requires the synthesis of new proteins, while WM can persist exclusively through the sustained neuronal activity of specific neuronal ensembles (32, 38, 400, 411). Certain brain areas such as the hippocampus, also seem to play a much greater role in LTM compared to WM, ultimately implying that although WM likely shares a degree of overlap with LTM, it is unlikely that these functions can be considered the same (412). WM duration, along with capacity, thus represents key variables for demarcating the two memory mechanisms.

WM Duration for different modalities of stimuli

The limits of WM duration are exemplified by the progressive decrease in performance over time, regardless of modality. Performance is thought to be dependent on the activity of the brain areas maintaining each type of information. In auditory tasks, for example, the maintenance of single tones was correlated with sustained activity in the interconnected areas of the auditory cortex, the inferior frontal gyrus, and the hippocampus, among which the former two were also able to distinguish the maintained tone (407). Likewise, when tactile WM information was being maintained, the application of transcranial magnetic stimulation was able to demonstrate a causal role for the middle frontal gyrus, but only for areas that possessed anatomical connections with the primary somatosensory cortex (413, 414). Importantly however, certain WM task contexts can result in overlap between otherwise modality-specific areas, even if the actual task stimuli are exclusively unimodal (415, 416). An excellent example can be observed with the maintenance of phonological information, such as silent lip movements—a purely visual stimulus that is also able to activate auditory areas when the lip movements resemble the formation of words (416). Certain tasks involving spatial frequency, speed, direction and motion coherence, seemed to undergo only limited decreases in performance when the required duration was extended by a significant degree (417-419). However, these opposition points have been criticized for failing to account for the special status of spatial information in visual WM, as well as the possibility of non-articulatory rehearsal via attention, which has been experimentally shown to supplement recall (382, 420).

The Role of Rehearsal in WM Duration

The maintenance of information in WM is greatly enhanced when conscious attention is devoted to the process, and various forms of rehearsal may be applied to this end. In particular, maintenance rehearsal is typically defined as the repetition of information, with a common example being the recitation of a phone number until the call is dialed, after which point, the number is rapidly forgotten (24, 421). Maintenance rehearsal results in similar neural activity as the original encoding of information when examined via EEG (422). Moreover, it seems possible that rehearsal may be “offloaded” to brain areas beyond the PFC, such as the modality specific areas that were previously noted. These critical insights thus reveal a possible key for explaining the neural mechanisms of rehearsal, and by extension, determining the degree to which this ability may ultimately be responsible for WM duration. On an extreme view, this has even led to the question of whether WM duration actually exists, or instead, if the phenomenon that we refer to as WM duration is

actually a product of rehearsal, due to the fact that objects seem to persist indefinitely in WM as long as the rehearsal continues to be maintained (423).

The rehearsal explanation for WM duration is further supported by the fact that the loss of WM content over time does not seem to be an all or nothing event, but instead, can be mitigated by devoting additional cognitive resources to specific objects, as demonstrated by the classic Sperling et al. experiment. Specifically, individuals were briefly (between 15-500 ms) presented with a 3x4 grid of random letters, and then asked to report the letters at each location (424). The average results for this first experimental condition ranged between four or five accurate letters. However, in a second experimental condition, participants were only required to report a single row of letters, indicated by a tone that would sound 300 ms after the array disappeared. The average results for this second experimental condition ranged between three or four accurate letters, double the per-row average of the original experimental condition, which implied that neural representations of most letters were still available after 300 ms, and that they could survive longer once attention consciously directed to exclusively maintaining the prompted row, while allowing all others to rapidly fade (424).

Decay, Interference, and Drift Models of WM Duration

The cause of the WM duration limits has been debated. It has been theorized that either spontaneous decay of sustained neuronal activity, interference and eventual displacement by new items, or gradual drifting in the activity of the neural ensembles that represent each remembered object may be the mechanism of eventual loss of the information maintained (38, 391, 425). For example, several studies found that distractors cause greater decreases in WM performance when they match the modality or category of the task relevant information (426-428). The implication is that the greater overlap between the task relevant information and the distractor leads to greater interference, and subsequently, greater reductions in performance. In turn, this idea would imply that WM is maintained as long as the relevant coding remains intact, however, the maintenance requires increasing cognitive resources over time. As a result, old content is gradually replaced by the new, unless attention is applied through abilities like rehearsal, to reserve cognitive resources on specific objects over time.

A challenge in neurophysiological studies that could chart neuronal activity as a function of visual working memory duration has been that it is difficult to avoid eye movements (and blinks) over prolonged delays, and neural activity driven by eye movements and retinal input may confound the interpretation of activity relating to the contents of working memory. However, some studies have succeeded (e.g. by allowing temporary timeouts in fixation) and they provided little evidence for decay of persistent activity during the duration of the trial (429). Neurophysiological studies of auditory and tactile working memory which are not subject to this limitation have similarly provided little evidence of decay, at least for intervals of ~6 seconds (339, 356). It is only at delay intervals of up to 15 seconds that activity of some individual prefrontal neurons begins to shut down, though other neurons are able to maintain persistent activity (430). Even in these experiments, LFP recordings

maintain their tuning for the preceding stimulus all the way up to the longest intervals tested, of approximately 15 seconds (431).

In the context of the bump attractor model, strong evidence exists that inaccuracies in recall are the effects of drift rather than decay of activity in the network (Fig. 11), at least for short periods of delay, of 1-3 seconds. Although it is well established that the accuracy of recalling the location of a stimulus decreases monotonically as a function of time (432), this does not necessarily imply decay of persistent activity. Instead, longer delays allow the bump of activity to drift and deviate more from the remembered location, on average. Critical predictions of the alternative, decay model, are not borne by the experimental data (55).

Working Memory across Species

Working memory is not exclusive to primates. Commonalities and differences between the WM mechanisms across and between different species reveals the roles of various neural mechanisms and their relative importance to various WM functions. For example, in virtually all animal models, the development of WM seems to be intertwined with additional cognitive systems, including the motor, sensory and attentional functions (433, 434). Substantial challenges also exist. In the broader sense, the incompleteness and only vague definition of psychological terms used to explain the neural basis of cognitive function comes into sharper relief the further away we wander in the evolutionary scale (10). The homology or even existence of areas such as the prefrontal cortex across species is also debated (435). Here we will review the neurophysiology of WM across the rodent and avian models, focusing particularly on the varying capabilities of each taxonomic class, the role of persistent activity, and the cortical areas involved in it.

Rodent Models

The rodent model of WM, including mice and rats, is one of the most commonly applied models in current research. Rodents can readily be trained to perform working memory tasks, including spatial working memory tasks, based on the radial maze, delayed response and delayed alternation tasks. Moreover, working memory tasks based on other sensory modalities, including olfactory, auditory, and tactile working memory have also been demonstrated. These paradigms have been used to address some of the same questions that have been raised in primate and human research, including the neural basis of working memory, where in the brain it is localized, the local and long-distance circuits that are essential for its maintenance and the limits of its capacity and duration. Some of the same debates are also present in the rodent literature.

Persistent activity has been reported across working memory tasks. Initial reports documented persistent activity in delayed response tasks (436), which might have been confounded with motor preparation, but newer studies have demonstrated persistent activity in a variety of tasks representing remembered features of stimuli across various modalities, e.g. auditory or tactile frequency (437, 438). In contrast to the primate persistent activity, however, individual rodent cortex neurons are not active through the entire delay period of WM tasks, but different groups of neurons become active for brief periods of time (Fig. 12),

so that their activity “tiles” the entire delay period (439). Information may be encoded by neurons that do not exhibit persistent activity, and by different neurons that become active during the course of learning a behavioral task (440).

As has been speculated in the primate model, different types of interneurons play a specialized role in the circuits that maintain working Memory. Optogenetic inhibition of different types of interneurons produce different results in WM performance: inactivation of parvalbumin and somatostatin interneurons degrade working memory performance, whereas inactivation of VIP neurons may improve it (438). Similar to the primate brain, persistent activity generated in the frontal lobe is also dependent upon long-range circuits which includes the mediodorsal thalamic nucleus (441). MD inactivation degrades performance and activity in the frontal lobe.

As has also been discussed for the primate brain, working memory activity is not limited only to the rodent prefrontal cortex. Areas have also been implicated as major factors in WM, including the posterior parietal cortex, the frontal orienting fields, the hippocampus, and the motor cortex—particularly in the anterolateral area (48, 436, 442, 443). These areas appear to be specialized in WM, and the frontal orienting fields, for example, have demonstrated selectivity to orientation in 37% of their population, with performance decreasing when these areas (48) are deactivated (436). Similarly, the hippocampus—typically associated with LTM—has demonstrated significant selectivity for spatial WM tasks.

Important differences in functions of prefrontal areas between rodent and primate species have also been noted. For example, lesions of the medial rodent PFC were reported to impair attention but not WM (444) and to not affect task performance in a two-choice serial reaction time task, though other studies reported impairments following medial PFC lesions when the number of possible targets in the serial reaction time task are increased (445, 446). The prospect of reconciling the conflicting results of these prior investigations is complicated by the lack of consensus regarding the anatomy of the rodent PFC and the question of whether the rodent and human PFC are truly homologous (435, 447). In particular, a recent meta-analysis has suggested that although medial prefrontal and cingulate areas are present in the rodent prefrontal cortex, the lateral prefrontal cortex, which is the focal area of persistent activity generation, has no clear homolog in the rodent brain (435). It is therefore possible that unique WM functions evolved along with unique PFC areas in the primate.

Avian Models

Many avian species—particularly corvids—have demonstrated high levels of WM, including a capacity for storing up to four distinct items and the ability to perform highly complex tasks (448, 449). However, they represent a unique model in WM, as their brains are highly dissimilar to the mammalian brain, particularly in terms of the structures typically associated with higher functions such as WW (450, 451). Notably, the avian brain seems to lack a PFC, neocortex or any laminated component within its telencephalon, despite possessing a similar structure of modular networks (451-454). This implies convergent evolution, with these two

taxonomic classes taking separate developmental paths to independently pursue the same functions (448, 455, 456).

The caudolateral nidopallium has been identified as an area important for working memory, demonstrating persistent activity when information is maintained in WM through a delay period in a wide variety of tasks. For example, during a spatial delayed-response task, at least 30% of the caudolateral corvid nidopallium demonstrated selectivity to both visual presentation and the maintenance of items during the delay period, thus resulting in decodable representations that remained stable throughout the course of the task (457). Persistent discharges have thus been described that can maintain visual spatial location and object identity, auditory stimuli, numerical quantity, and abstract task rules (458-460). The parallel between the caudolateral nidopallium and the PFC is further supported by the fact that both areas are central hubs for motor and sensory activity, the development of which are known to be intertwined with WM (433, 434, 461). The direct causality of the caudolateral nidopallium in WM is suggested by a study where the controlled lesioning of the caudolateral nidopallium was found to generate WM deficits that were proportional to the size of the lesion (462). Recently however, these results have been questioned in certain avian species, such as pigeons, due to the fact that persistent neural activity has not been observed in all studies of the caudolateral nidopallium, and may be modulated by factors, such as the presence of a task reward (463, 464). This may imply that the caudolateral nidopallium may only be the seat of certain aspects of WM. However, the wide range of cognitive abilities between different avian species may also result in some species simply lacking the ability to effectively conduct certain types of WM tasks.

Working Memory Development

Humans and nonhuman primates experience a protracted period of cognitive development that parallels the maturation of the prefrontal cortex (465-469). Working memory ability, in particular, undergoes significant improvements that are one of the hallmarks of human cognitive development (470-475). A monotonic age-based increase in the performance of multiple types of working memory tasks has been reported, including improvement in the mean accuracy and reaction time of memory-guided saccade tasks (470, 471, 473). Tasks that assess verbal and visuo-spatial working memory and executive control show parallel developmental profiles (470, 476). Developmental improvements in WM are not only characterized by an increase in correct trials, but also by a decrease in performance variability from trial to trial, which mirrors decreases in the variability of whole brain states of task-related activity (477). On the other hand, neurodevelopmental disorders that manifest in childhood and late adolescence, such as ADHD and schizophrenia, result in corresponding deficits of working memory (478-482).

Age-related performance increases observed for working memory tasks are associated with structural changes due to white and gray-matter maturation in the PFC and its connections with other cortical regions in humans (483, 484). Human fMRI studies have also shown that patterns of brain activation associated with the performance of working memory tasks undergo distinct changes between childhood and adulthood, typically involving increased

activation of frontal and parietal areas, but also more effective deactivation of other areas, such as the default mode network (485-489).

The nonhuman primate model has been particularly fruitful for understanding working memory maturation. The male rhesus monkey (*Macaca mulatta*) enters puberty at approximately 3.5 years of age and reaches full sexual maturity at 5 (490, 491), roughly equivalent to human ages of 11 and 16 years, respectively thus making this a particularly well-suited model to inform human neurocognitive development. Similar to the human pattern of development, the monkey prefrontal cortex continues to mature during adolescence and into early adulthood (492-494), establishing connections with a network of brain areas, similar to that of the human PFC. Biochemical and anatomical changes have also been characterized within the monkey prefrontal cortex from pre-puberty to adulthood, including maturation of interneuron morphology and connections (493, 495).

Behavioral performance and neural activity in working memory tasks change markedly around the time of puberty (Fig. 13), a developmental event associated with the release of sex hormones and significant neurological change (77, 496). Performance of working memory tasks is subtly but significantly higher in adult monkeys compared to adolescent monkeys that have entered puberty. The improvement involves both increases in mean performance and decreases in variability, similar to the improvements observed in humans between these two developmental stages (497). Most importantly, persistent activity during the delay intervals of working memory tasks is higher in adult animals than adolescent ones. Even when comparing persistent activity from adolescent and adult monkeys obtained in sessions equated for performance, the adult prefrontal cortex is better able to generate persistent activity (496). Furthermore, the activity that follows presentation of distractors—stimuli that the monkeys have been instructed to ignore—is substantially lower in adulthood, suggesting that the adult prefrontal cortex is better able to filter distracting stimuli during working memory. These changes refer specifically to persistent activity. Responses to visual stimuli themselves are relatively unchanged, indicating that stimulus-driven neuronal responses are fully mature by puberty (496).

The observed changes in neural activity during childhood, adolescence, and adulthood suggest plasticity in the circuit that generates persistent discharges. An example of this may be observed in the local-circuit differences between adolescent and adult monkeys. Recordings from nearby pairs of neurons (within 0.5 – 1 mm from each other) reveal a higher percentage of spikes in adult than adolescent monkeys that are generated in near synchrony, within 2-3 ms of each other (498). This increase in “zero-lag synchronization” with age is the result of weakened inhibitory interactions in adulthood, either as a result of pruning of inhibitory connections, or decreases in the connectivity strength of pyramidal neurons onto interneurons, which lessens the net output of inhibitory connections as the prefrontal cortex matures (499). Interestingly, a decrease in zero-lag synchrony of prefrontal neurons has been recently implicated in an animal model of schizophrenia (185), a condition that, among other pathological symptoms, compromises working memory.

Changes in neuronal activity and connectivity that are observed in aged monkeys somewhat reverse the changes that are seen between adolescence and adulthood. Most importantly,

persistent activity during working memory tasks is substantially lower in old monkeys, though in this case too, firing rate during the stimulus presentation differed little between age groups (79). Aged animals exhibit elevated cyclic-AMP (cAMP) signaling, which reduces persistent activity by opening Hyperpolarization-activated Cyclic Nucleotide-gated channels (HCN - nonselective voltage-gated cation channels), and KCNQ, (Potassium voltage-gated channels). Persistent activity can be partially restored to more youthful levels by inhibiting cAMP signaling, or by blocking HCN or KCNQ channels (79).

Working Memory Training

Working memory training may induce significant performance improvements at all stages of life, including adulthood (38). Computerized training has resulted in expanded WM capacity, as well as clinical improvements for patients of stroke, ADHD, schizophrenia (500-503) as well as aged adults (504). However, there is a heated debate in regards to if and how these improvements may generalize (“transfer”) to tasks that were not part of the training, including more complex day to day activities (505, 506). The view against the benefits of training argues that improved performance levels are actually the result of task-specific skills instead of genuinely expanded abilities (507). Imaging studies have also produced conflicting results about the effects of training, with some studies suggesting increased activity in executive brain areas (503, 508-512) and others suggesting decreased activity instead (513-516). Thus increased activity has been suggested to reflect specialized selectivity, while decreases in activity are likewise suggested as an improvement in efficiency (38). To resolve these debates, it is again necessary to understand changes at the level of neural activity as a result of training.

Initial Working Memory Training

The appearance of persistent activity seems to be ubiquitous regardless of the task type, the subject’s prior training, or even whether they were required to perform a working memory task, at all. For example, persistent prefrontal activity has been observed in naïve monkeys that passively view stimuli without performing an active task, continuing even after the stimuli are no longer present (56, 314). This evidence—and the everyday ability to track our environment and recall information even when not explicitly prompted to do so ahead of time—would collectively argue against the view that working memory necessarily requires conscious effort. Thus, working memory seems to be an automatic function that is only subject to, but not dependent upon, conscious control. Importantly, the observed persistent activity in naïve monkeys is selective for specific properties of the stimuli, including spatial location, color, and shape (56, 314). There are also limits to the information that may be represented automatically. In particular, the decodable identity of a stimulus would not persist after the presentation of a new stimulus in the experiments discussed above, and the information of whether two successively presented stimuli were the same or not was largely absent in naïve animals (56, 64, 314).

Training of monkeys to perform a working memory task for the first time results in long lasting changes in neural activity (Fig. 14), including an overall increase in persistent activity, in addition to increases in the number of neurons that are activated after appearance

of stimuli (56, 66, 314, 517). These changes are evident even when the trained monkeys are tested with passive presentation of stimuli in the same fashion they did prior to training, and are predictably emphasized when they actively participate (314). Moreover, greater training seems to result in greater changes in activity that reflect the task performance levels at each point in time (73, 399).

Prefrontal subdivisions along the anterior-posterior and dorso-ventral axes of the PFC display varying degrees of plasticity after training, as already noted in the section of prefrontal areal specialization. Along the dorsal PFC, the highest levels plasticity occur in more anterior areas, including the mid-, and anterior-dorsal areas (area 46). On the other hand, the posterior dorsal prefrontal areas exhibit limited training effects beyond a small increase in the percentage of posterior-dorsal neurons that are activated by stimuli (66, 95, 96). A similar gradient of plasticity runs along the dorso-ventral axis, with most plastic responses observed in the ventral PFC and least plastic in the more medial aspects of the dorsal PFC (95). Areas along the principal sulcus exhibit intermediate levels of plasticity in terms of increases in persistent discharges after WM training.

Decoding information about remembered stimuli depends not only on their mean firing rate, but also the trial-to-trial response variability, and the positive correlation of firing rates between neurons, all of which are affected by changes in plasticity (88, 518, 519). For example, training decreases the Fano factor of spike counts, a measure of variability, with the greatest decreases observed in neurons that exhibit persistent activity. This implies a refinement of neural selectivity and reliability of stimulus property representations. Similarly, the post training decrease in noise correlation—the spike-count correlation of persistent firing rates between pairs of neurons—also improves information decoding from simultaneously active neurons (519). Task training also alters the time course and dynamics of persistent activity (399, 520). For example, the firing rates of some neurons are known to “ramp up” or decrease during the course of the trial, in order to encode information dynamically at different time points (64, 147, 339). With training, these changes become more pronounced in trials compared to animals who were only assigned to passive fixation (399, 520).

Plasticity when learning to perform additional tasks

Subjects that have already trained to perform a cognitive task demonstrate that persistent activity can undergo rapid modification when they are exposed to a new task condition or type of stimulus (521). Individual neurons may begin to be activated by stimuli that they were initially unresponsive to over the course of a single training session, when the animals are learning to associate these particular stimuli with motor responses that earn reward, e.g. “look to the left when stimulus A is presented” (522). The ability of prefrontal neurons to rapidly represent different aspects of stimuli in persistent activity is also illustrated after training in tasks that require animals to remember one relevant stimulus feature and ignore others (120). Dual-task paradigms reveal that persistent activity can rapidly decrease and increase again over the course of a single trial, as the subject alternates their focus between the two tasks (523).

Training-induced increases in plasticity may also point to the specialization of underlying cellular and molecular mechanisms (524) between specific prefrontal subdivisions. For example, recent studies have demonstrated the systematic variation of plasticity markers between limbic and eulaminate areas, including the relative scarcity of Calcium/calmodulin-dependent protein kinase II (CaMKII) in area 46d compared to more anterior limbic areas (525). Markers of cortical stability, including intracortical myelin, perineuronal nets, and parvalbumin have demonstrated the opposite pattern. Changes in neuronal morphology, molecular profiles of the synaptic apparatus, and the influence of neuromodulator systems have also been implicated in long-term prefrontal plasticity (526, 527), and may differ between areas. Further mechanisms for post training plasticity may be connected to observed changes in short-term prefrontal synaptic plasticity, depression, and facilitation (528).

The Effects of Training on WM Capacity

Although WM capacity was previously thought to be a fixed quality in individuals, recent research has shown that training can actually enhance WM capacity—particularly in children and young adults—with significant improvements that extend beyond the training-specific tasks (487, 529, 530). This is important for assisting the rehabilitation of individuals with memory related disorders, such as ADHD, and also for improving the cognitive abilities of neurotypical individuals. Moreover, the modification of WM capacity highlights possible neural mechanisms that may mediate capacity.

Firing rate—Training in tasks that require the simultaneous storage of multiple stimuli can induce plastic changes in prefrontal activity, activating more neurons, decreasing baseline firing, and increasing the rate of persistent activity in order to ultimately improve WM capacity—at least in regard to the trained task (399). Training may also occasionally decrease persistent activity, and the reason for this is currently debated as an improvement in efficiency or strategy (38, 513-516). It is also possible that change detected in fMRI may be due to decreased baseline activity, particularly when activity is averaged over long periods. Both humans and monkeys are known to develop personalized strategies that assist their performance in complex tasks, e.g. suggestive of grouping of multiple stimuli in memory based on their geometric arrangement (126). These strategies are reflected in the altered prefrontal response selectivity and dynamics when comparing single and multi-stimuli tasks, even if the stimuli sets are the same between both task types (124, 399). In particular, a greater percentage of neural activity from individuals who displayed post-training improvements in multiple stimuli task performance could be explained by “condition-independent” components, not related to any remembered stimuli (399).

Dopamine—A variety of PET studies have investigated the role of dopamine in WM capacity before and after training, finding that changes in D1R density can be mapped to changes in WM capacity in an inverted U graph (531). This was reinforced by similar results that were observed when applying D1R antagonists (532). An additional study was conducted to assess D2R density, finding that there was increased dopamine release in individuals who were currently performing a memory task that could be linked to previously observed increases in BOLD activity after training (533). These insights are particularly

compelling in light of the abnormal dopamine release that is observed in ADHD and schizophrenia, among other memory-related disorders, as a possible explanation for the associated WM deficits (534, 535).

Functional connectivity—White matter volume and structure has long been associated with WM capacity (536-539). Structural connectivity—particularly from the PFC to its afferent cortical areas—has previously been correlated with WM capacity during development (398, 539, 540). In fact, the role of functional connectivity does not seem to be limited to development alone. In a study where adults performed n-back WM training, improved performance was correlated with increases in functional connectivity, with the greatest increases occurring between the parietal cortex and a superior frontal region, as well as between the parietal cortex and the dorsolateral PFC (540). In the above study on functional connectivity—and many others that attempt to confront the effects of training on WM capacity—it is worth noting that improved task performance does not necessarily equate to improved capacity. Instead, the observed improvement may be due to improved WM processing or duration, among other factors.

Arguably the most dramatic effects of WM capacity training can be observed in the case of professional mnemonists, who can rapidly store exceptional large quantities of items—usually letters or numerical digits—in their WM. For example, an individual known as SF was presented with random auditory digits at a rate of 1 second per digit, and, after 250 hours of practice, was able to raise his recall limit from 7 digits, to 80 (541). These results were later replicated by another subject, BB, though BB only received 86 hours of practice and required 5 seconds to store each digit in his WM to achieve maximum capacity (542). Notably, nearly all professional mnemonists are known to apply various encoding strategies, such as the famous method of loci, where objects are associated with familiar locations, after which the objects may be recalled by imaging the locations (543). These encoding strategies may therefore serve as an excellent method for demonstrating the effects of training on WM capacity in a clearly observable manner. However, the relative lack of research on the topic of encoding strategies may bring their use into question, and among followers of the previously discussed discrete units theory for WM capacity, there is the possibility that encoding strategies are not genuinely increasing WM capacity, but instead, simply serve to more effectively organize chunks of information into each discrete “unit” of an individual’s existing WM capacity. Alternatively, it is also possible that the encoding strategies may serve to rapidly store information in long term memory, allowing the recall of LTM to supplant WM without requiring an increase in capacity. Ultimately, these questions can only be resolved by bringing the encoding strategies of mnemonists back to the forefront of research and continuing the work that was started by the prior generation of neuroscientists.

Conclusions

Recent research has been able to provide a comprehensive link between the activity of neurons in the prefrontal cortex and other brain areas with a range of psychological phenomena related to working memory. The bump attractor model of visuospatial working memory provides a basic framework that links spiking activity with behavior. Synaptic mechanisms allow representation of latent information even in time intervals of no overt

spiking activity, but they influence behavior inasmuch as they influence the generation of spiking activity. Limits of capacity and duration can be traced back to the properties of this model, as can improvement of working memory during childhood development and through adult training. Flexible behavior requires considerable neuronal plasticity, however some specialization of circuits mediating different types of information within the prefrontal cortex and between its connections is still evident.

A number of unanswered questions remain. The role of rhythmicity in working memory is still unresolved. Rhythmic discharges and local field potentials can be readily observed in working memory paradigms. The extent to which they represent information beyond what firing rate represents and whether they influence behavior is unclear. The circuits that maintain memory for objects have been more difficult to investigate. Whether fundamentally different mechanisms participate in object than spatial working memory maintenance is an open question. The intrinsic organization of the prefrontal cortex also remains unanswered. Hints of local organization suggest that the representation of information is not entirely random across the cortical surface. Future studies addressing these questions will help tie cognitive phenomena to the activity of cortical neurons.

Acknowledgments

Research reported in this paper was supported by the National Eye Institute and National Institute of Mental Health of the National Institutes of Health under award number R01 EY017077 and MH116675.

References

1. Baddeley A (2012) Working memory: theories, models, and controversies. *Annu Rev Psychol* 63:1–29. [PubMed: 21961947]
2. Baddeley A (1992) Working memory. *Science* 255(5044):556–559. [PubMed: 1736359]
3. Morris RG & Baddeley AD (1988) Primary and working memory functioning in Alzheimer-type dementia. *J Clin Exp Neuropsychol* 10(2):279–296. [PubMed: 3280591]
4. Martinussen R, Hayden J, Hogg-Johnson S, & Tannock R (2005) A meta-analysis of working memory impairments in children with attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 44(4):377–384. [PubMed: 15782085]
5. Forbes NF, Carrick LA, McIntosh AM, & Lawrie SM (2009) Working memory in schizophrenia: a meta-analysis. *Psychol Med* 39(6):889–905. [PubMed: 18945379]
6. Rossi AF, Bichot NP, Desimone R, & Ungerleider LG (2007) Top down attentional deficits in macaques with lesions of lateral prefrontal cortex. *J Neurosci* 27(42):11306–11314. [PubMed: 17942725]
7. Curtis CE & D'Esposito M (2004) The effects of prefrontal lesions on working memory performance and theory. *Cogn Affect Behav Neurosci* 4(4):528–539. [PubMed: 15849895]
8. Suchow JW, Fougny D, Brady TF, & Alvarez GA (2014) Terms of the debate on the format and structure of visual memory. *Atten Percept Psychophys* 76(7):2071–2079. [PubMed: 24894917]
9. Miller GA (1960) *Plans and the structure of behavior* (Holt, New York,) p 226 p.
10. Buzsaki G (2020) The Brain-Cognitive Behavior Problem: A Retrospective. *eNeuro* 7(4).
11. Poeppel D & Adolphi F (2020) Against the Epistemological Primacy of the Hardware: The Brain from Inside Out, Turned Upside Down. *eNeuro* 7(4).
12. Awh E, Vogel EK, & Oh SH (2006) Interactions between attention and working memory. *Neuroscience* 139(1):201–208. [PubMed: 16324792]
13. Fukuda K, Awh E, & Vogel EK (2010) Discrete capacity limits in visual working memory. *Curr Opin Neurobiol* 20(2):177–182. [PubMed: 20362427]

14. Oberauer K (2019) Working Memory and Attention - A Conceptual Analysis and Review. *J Cogn* 2(1):36. [PubMed: 31517246]
15. Malmberg KJ, Raaijmakers JGW, & Shiffrin RM (2019) 50 years of research sparked by Atkinson and Shiffrin (1968). *Mem Cognit* 47(4):561–574.
16. Zaksas D & Pasternak T (2006) Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. *J Neurosci* 26(45):11726–11742. [PubMed: 17093094]
17. Atkinson AL, et al. (2018) Are there multiple ways to direct attention in working memory? *Ann N Y Acad Sci* 1424(1):115–126. [PubMed: 29635690]
18. Plancher G & Barrouillet P (2020) On some of the main criticisms of the modal model: Reappraisal from a TBRS perspective. *Mem Cognit* 48(3):455–468.
19. Baddeley A (2000) The episodic buffer: a new component of working memory? *Trends Cogn Sci* 4(11):417–423. [PubMed: 11058819]
20. Cowan N (2001) The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci* 24(1):87–114; discussion 114–185. [PubMed: 11515286]
21. Eng HY, Chen D, & Jiang Y (2005) Visual working memory for simple and complex visual stimuli. *Psychon Bull Rev* 12(6):1127–1133. [PubMed: 16615339]
22. Baddeley AD & Andrade J (2000) Working memory and the vividness of imagery. *J Exp Psychol Gen* 129(1):126–145. [PubMed: 10756490]
23. Norris D (2019) Even an activated long-term memory system still needs a separate short-term store: A reply to Cowan (2019). *Psychol Bull* 145(8):848–853. [PubMed: 31328942]
24. Gardiner JM, Gawlik B, & Richardson-Klavehn A (1994) Maintenance rehearsal affects knowing, not remembering; elaborative rehearsal affects remembering, not knowing. *Psychon Bull Rev* 1(1): 107–110. [PubMed: 24203417]
25. Rubinstein JS, Meyer DE, & Evans JE (2001) Executive control of cognitive processes in task switching. *J Exp Psychol Hum Percept Perform* 27(4):763–797. [PubMed: 11518143]
26. Sandhu R & Dyson BJ (2013) Modality and task switching interactions using bi-modal and bivalent stimuli. *Brain Cogn* 82(1):90–99. [PubMed: 23524240]
27. Miyake A, et al. (2000) The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cogn Psychol* 41(1):49–100. [PubMed: 10945922]
28. Whitney P, Arnett PA, Driver A, & Budd D (2001) Measuring central executive functioning: what's in a reading span? *Brain Cogn* 45(1):1–14. [PubMed: 11161358]
29. Parkin AJ (1998) The central executive does not exist. *J Int Neuropsychol Soc* 4(5):518–522. [PubMed: 9745241]
30. D'Esposito M & Postle BR (2015) The cognitive neuroscience of working memory. *Annu Rev Psychol* 66:115–142. [PubMed: 25251486]
31. Riley MR & Constantinidis C (2016) Role of prefrontal persistent activity in working memory. *Front Syst Neurosci* 9:181. [PubMed: 26778980]
32. Rigotti M, et al. (2013) The importance of mixed selectivity in complex cognitive tasks. *Nature* 497(7451):585–590. [PubMed: 23685452]
33. Todd JJ & Marois R (2004) Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature* 428(6984):751–754. [PubMed: 15085133]
34. Kuo BC, Stokes MG, Murray AM, & Nobre AC (2014) Attention biases visual activity in visual short-term memory. *J Cogn Neurosci* 26(7):1377–1389. [PubMed: 24456394]
35. Sprague TC, Ester EF, & Serences JT (2016) Restoring Latent Visual Working Memory Representations in Human Cortex. *Neuron* 91(3):694–707. [PubMed: 27497224]
36. Sprague TC, Ester EF, & Serences JT (2014) Reconstructions of information in visual spatial working memory degrade with memory load. *Curr Biol* 24(18):2174–2180. [PubMed: 25201683]
37. Deubel H & Schneider WX (1996) Saccade target selection and object recognition: evidence for a common attentional mechanism. *Vision Res* 36(12):1827–1837. [PubMed: 8759451]
38. Constantinidis C & Klingberg T (2016) The neuroscience of working memory capacity and training. *Nat Rev Neurosci* 17(7):438–449. [PubMed: 27225070]

39. Lebedev MA, Messinger A, Kralik JD, & Wise SP (2004) Representation of attended versus remembered locations in prefrontal cortex. *PLoS Biol* 2(11):e365. Epub 2004 Oct 2026. [PubMed: 15510225]
40. Soto D & Silvanto J (2014) Reappraising the relationship between working memory and conscious awareness. *Trends Cogn Sci* 18(10):520–525. [PubMed: 25070269]
41. Soto D, Mantyla T, & Silvanto J (2011) Working memory without consciousness. *Curr Biol* 21(22):R912–913. [PubMed: 22115455]
42. Chelazzi L, Miller EK, Duncan J, & Desimone R (1993) A neural basis for visual search in inferior temporal cortex. *Nature* 363(6427):345–347. [PubMed: 8497317]
43. Treue S & Maunsell JH (1996) Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* 382(6591):539–541. [PubMed: 8700227]
44. Leavitt ML, Mendoza-Halliday D, & Martinez-Trujillo JC (2017) Sustained Activity Encoding Working Memories: Not Fully Distributed. *Trends Neurosci* 40(6):328–346. [PubMed: 28515011]
45. Rose NS, et al. (2016) Reactivation of latent working memories with transcranial magnetic stimulation. *Science* 354(6316):1136–1139. [PubMed: 27934762]
46. Squire LR & Wixted JT (2011) The cognitive neuroscience of human memory since H.M. *Annu Rev Neurosci* 34:259–288. [PubMed: 21456960]
47. Ranganath C, Cohen MX, & Brozinsky CJ (2005) Working memory maintenance contributes to long-term memory formation: neural and behavioral evidence. *J Cogn Neurosci* 17(7):994–1010. [PubMed: 16102232]
48. Hampson RE, Simeral JD, & Deadwyler SA (1999) Distribution of spatial and nonspatial information in dorsal hippocampus. *Nature* 402(6762):610–614. [PubMed: 10604466]
49. Mongillo G, Barak O, & Tsodyks M (2008) Synaptic theory of working memory. *Science* 319(5869):1543–1546. [PubMed: 18339943]
50. Zhang WH & Williams ZM (2015) Frontal neurons modulate memory retrieval across widely varying temporal scales. *Learn Mem* 22(6):299–306. [PubMed: 25979992]
51. Fuster JM & Alexander GE (1971) Neuron activity related to short-term memory. *Science* 173(997):652–654. [PubMed: 4998337]
52. Funahashi S, Bruce CJ, & Goldman-Rakic PS (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology* 61(2):331–349. [PubMed: 2918358]
53. Kaminski J, et al. (2017) Persistently active neurons in human medial frontal and medial temporal lobe support working memory. *Nat Neurosci* 20(4):590–601. [PubMed: 28218914]
54. Qi XL, Zhou X, & Constantinidis C (2015) Neurophysiological Mechanisms of Working Memory: Cortical Specialization & Plasticity. *Attention and Performance XXV*, eds Jolicoeur P, Lefebvre C, & Martinez-Trujillo JC (Academic Press, London), pp 171–186.
55. Wimmer K, Nykamp DQ, Constantinidis C, & Compte A (2014) Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nat Neurosci* 17(3):431–439. [PubMed: 24487232]
56. Meyer T, Qi XL, & Constantinidis C (2007) Persistent discharges in the prefrontal cortex of monkeys naive to working memory tasks. *Cereb Cortex* 17 Suppl 1:i70–76. [PubMed: 17726005]
57. Riley MR & Constantinidis C (2015) Role of Prefrontal Persistent Activity in Working Memory. *Front Syst Neurosci* 9:181. [PubMed: 26778980]
58. Mishkin M, Ungerleider LG, & Macko KA (1983) Object vision and spatial vision: Two cortical pathways. *Trends in Neuroscience* 6:414–417.
59. Wilson FA, Scalaidhe SP, & Goldman-Rakic PS (1993) Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science* 260(5116):1955–1958. [PubMed: 8316836]
60. Rao SG, Williams GV, & Goldman-Rakic PS (1999) Isodirectional tuning of adjacent interneurons and pyramidal cells during working memory: evidence for microcolumnar organization in PFC. *Journal of Neurophysiology* 81(4):1903–1916. [PubMed: 10200225]

61. Constantinidis C, Franowicz MN, & Goldman-Rakic PS (2001) Coding specificity in cortical microcircuits: a multiple electrode analysis of primate prefrontal cortex. *Journal of Neuroscience* 21(10):3646–3655. [PubMed: 11331394]
62. Niki H (1974) Prefrontal unit activity during delayed alternation in the monkey. I. Relation to direction of response. *Brain Research* 68(2):185–196. [PubMed: 4363757]
63. Kubota K & Niki H (1971) Prefrontal cortical unit activity and delayed alternation performance in monkeys. *Journal of Neurophysiology* 34(3):337–347. [PubMed: 4997822]
64. Meyers EM, Qi XL, & Constantinidis C (2012) Incorporation of new information into prefrontal cortical activity after learning working memory tasks. *Proc Natl Acad Sci U S A* 109(12):4651–4656. [PubMed: 22392988]
65. Qi XL & Constantinidis C (2013) Neural changes after training to perform cognitive tasks. *Behav Brain Res* 241:235–243. [PubMed: 23261872]
66. Riley MR, Qi XL, Zhou X, & Constantinidis C (2018) Anterior-posterior gradient of plasticity in primate prefrontal cortex. *Nat Commun* 9(1):3790. [PubMed: 30224705]
67. Leavitt ML, Pieper F, Sachs AJ, & Martinez-Trujillo JC (2017) A Quadratic Bias in Prefrontal Representation of Visual-Mnemonic Space. *Cereb Cortex*:1–17. [PubMed: 28365777]
68. Lundqvist M, Herman P, & Miller EK (2018) Working Memory: Delay Activity, Yes! Persistent Activity? Maybe Not. *J Neurosci* 38(32):7013–7019. [PubMed: 30089640]
69. Markowitz DA, Curtis CE, & Pesaran B (2015) Multiple component networks support working memory in prefrontal cortex. *Proc Natl Acad Sci U S A* 112(35):11084–11089. [PubMed: 26283366]
70. Funahashi S, Chafee MV, & Goldman-Rakic PS (1993) Prefrontal neuronal activity in rhesus monkeys performing a delayed anti-saccade task. *Nature* 365(6448):753–756. [PubMed: 8413653]
71. Takeda K & Funahashi S (2002) Prefrontal task-related activity representing visual cue location or saccade direction in spatial working memory tasks. *J Neurophysiol* 87(1):567–588. [PubMed: 11784772]
72. Qi XL, et al. (2010) Comparison of neural activity related to working memory in primate dorsolateral prefrontal and posterior parietal cortex. *Front Syst Neurosci* 4:12. [PubMed: 20514341]
73. Qi XL, Meyer T, Stanford TR, & Constantinidis C (2011) Changes in Prefrontal Neuronal Activity after Learning to Perform a Spatial Working Memory Task. *Cerebral Cortex* 21:2722–2732. [PubMed: 21527786]
74. Goodwin SJ, Blackman RK, Sakellaridi S, & Chafee MV (2012) Executive control over cognition: stronger and earlier rule-based modulation of spatial category signals in prefrontal cortex relative to parietal cortex. *J Neurosci* 32(10):3499–3515. [PubMed: 22399773]
75. Masse NY, Hodnefield JM, & Freedman DJ (2017) Mnemonic Encoding and Cortical Organization in Parietal and Prefrontal Cortices. *J Neurosci* 37(25):6098–6112. [PubMed: 28539423]
76. Chafee MV & Goldman-Rakic PS (2000) Inactivation of Parietal and Prefrontal Cortex Reveals Interdependence of Neural Activity During Memory Guided-Saccades. *Journal of Neurophysiology* 83:1550–1566. [PubMed: 10712479]
77. Zhou X, et al. (2013) Working Memory Performance and Neural Activity in the Prefrontal Cortex of Peri-pubertal Monkeys. *J Neurophysiol* 110:2648–2660. [PubMed: 24047904]
78. Constantinidis C, Franowicz MN, & Goldman-Rakic PS (2001) The sensory nature of mnemonic representation in the primate prefrontal cortex. *Nature Neuroscience* 4(3):311–316. [PubMed: 11224549]
79. Wang M, et al. (2011) Neuronal basis of age-related working memory decline. *Nature* 476(7359):210–213. [PubMed: 21796118]
80. Mendoza-Halliday D, Torres S, & Martinez-Trujillo JC (2014) Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nat Neurosci* 17(9):1255–1262. [PubMed: 25108910]
81. Seung HS, Lee DD, Reis BY, & Tank DW (2000) Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron* 26(1):259–271. [PubMed: 10798409]

82. Lundqvist M, et al. (2016) Gamma and Beta Bursts Underlie Working Memory. *Neuron* 90(1):152–164. [PubMed: 26996084]
83. Hussar CR & Pasternak T (2012) Memory-guided sensory comparisons in the prefrontal cortex: contribution of putative pyramidal cells and interneurons. *J Neurosci* 32(8):2747–2761. [PubMed: 22357858]
84. Hussar CR & Pasternak T (2013) Common rules guide comparisons of speed and direction of motion in the dorsolateral prefrontal cortex. *J Neurosci* 33(3):972–986. [PubMed: 23325236]
85. Compte A, et al. (2003) Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. *J Neurophysiol* 28:3441–3454.
86. Li D, Constantinidis C, & Murray JD (2020) Trial-to-trial variability of spiking delay activity in prefrontal cortex constrains burst-coding models of working memory. *bioRxiv* 10.1101/2021.01.30.428962.
87. Chang MH, Armstrong KM, & Moore T (2012) Dissociation of response variability from firing rate effects in frontal eye field neurons during visual stimulation, working memory, and attention. *J Neurosci* 32(6):2204–2216. [PubMed: 22323732]
88. Qi XL & Constantinidis C (2012) Variability of prefrontal neuronal discharges before and after training in a working memory task. *PLoS ONE* 7(7):e41053. [PubMed: 22848426]
89. Hussar C & Pasternak T (2010) Trial-to-trial variability of the prefrontal neurons reveals the nature of their engagement in a motion discrimination task. *Proc Natl Acad Sci U S A* 107(50):21842–21847. [PubMed: 21098286]
90. Murray JD, et al. (2017) Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proc Natl Acad Sci U S A* 114(2):394–399. [PubMed: 28028221]
91. Machens CK, Romo R, & Brody CD (2010) Functional, But Not Anatomical, Separation of "What" and "When" in Prefrontal Cortex. *Journal of Neuroscience* 30(1):350–360. [PubMed: 20053916]
92. Druckmann S & Chklovskii DB (2012) Neuronal circuits underlying persistent representations despite time varying activity. *Curr Biol* 22(22):2095–2103. [PubMed: 23084992]
93. Rao SC, Rainer G, & Miller EK (1997) Integration of what and where in the primate prefrontal cortex. *Science* 276(5313):821–824. [PubMed: 9115211]
94. Wilson FA, Scalaidhe SP, & Goldman-Rakic PS (1993) Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science* 260(5116):1955–1958. [PubMed: 8316836]
95. Meyer T, Qi XL, Stanford TR, & Constantinidis C (2011) Stimulus selectivity in dorsal and ventral prefrontal cortex after training in working memory tasks. *J Neurosci* 31(17):6266–6276. [PubMed: 21525266]
96. Constantinidis C & Qi XL (2018) Representation of Spatial and Feature Information in the Monkey Dorsal and Ventral Prefrontal Cortex. *Front Integr Neurosci* 12:31. [PubMed: 30131679]
97. Hoshi E, Shima K, & Tanji J (1998) Task-dependent selectivity of movement-related neuronal activity in the primate prefrontal cortex. *Journal of Neurophysiology* 80(6):3392–3397. [PubMed: 9862940]
98. Sakagami M, et al. (2001) A code for behavioral inhibition on the basis of color, but not motion, in ventrolateral prefrontal cortex of macaque monkey. *J Neurosci* 21(13):4801–4808. [PubMed: 11425907]
99. Quintana J, Yajeya J, & Fuster JM (1988) Prefrontal representation of stimulus attributes during delay tasks. I. Unit activity in cross-temporal integration of sensory and sensory-motor information. *Brain Research* 474(2):211–221. [PubMed: 3208130]
100. Inoue M & Mikami A (2006) Prefrontal activity during serial probe reproduction task: encoding, mnemonic, and retrieval processes. *J Neurophysiol* 95(2):1008–1041. [PubMed: 16207786]
101. Genovesio A, Tsujimoto S, & Wise SP (2009) Feature- and order-based timing representations in the frontal cortex. *Neuron* 63(2):254–266. [PubMed: 19640483]
102. Averbeck BB, Chafee MV, Crowe DA, & Georgopoulos AP (2003) Neural activity in prefrontal cortex during copying geometrical shapes. Single cells encode shape, sequence, and metric parameters. *Exp Brain Res* 150(2):127–141. [PubMed: 12669170]

103. Roy JE, Buschman TJ, & Miller EK (2014) PFC neurons reflect categorical decisions about ambiguous stimuli. *J Cogn Neurosci* 26(6):1283–1291. [PubMed: 24405188]
104. Freedman DJ, Riesenhuber M, Poggio T, & Miller EK (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291(5502):312–316. [PubMed: 11209083]
105. O Scalaidhe SP, Wilson FA, & Goldman-Rakic PS (1999) Face-selective neurons during passive viewing and working memory performance of rhesus monkeys: evidence for intrinsic specialization of neuronal coding. *Cerebral Cortex* 9(5):459–475. [PubMed: 10450891]
106. O Scalaidhe S, Wilson FA, & Goldman-Rakic PS (1997) Areal segregation of face-processing neurons in prefrontal cortex. *Science* 278(5340):1135–1138. [PubMed: 9353197]
107. Rainer G, Asaad WF, & Miller EK (1998) Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature* 393(6685):577–579. [PubMed: 9634233]
108. Rainer G & Miller EK (2000) Effects of visual experience on the representation of objects in the prefrontal cortex. *Neuron* 27(1):179–189. [PubMed: 10939341]
109. Miller EK, Erickson CA, & Desimone R (1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience* 16(16):5154–5167. [PubMed: 8756444]
110. White IM & Wise SP (1999) Rule-dependent neuronal activity in the prefrontal cortex. *Exp Brain Res* 126(3):315–335. [PubMed: 10382618]
111. Wallis JD, Anderson KC, & Miller EK (2001) Single neurons in prefrontal cortex encode abstract rules. *Nature* 411(6840):953–956. [PubMed: 11418860]
112. Shima K, Isoda M, Mushiake H, & Tanji J (2007) Categorization of behavioural sequences in the prefrontal cortex. *Nature* 445(7125):315–318. [PubMed: 17183266]
113. Nieder A, Freedman DJ, & Miller EK (2002) Representation of the quantity of visual items in the primate prefrontal cortex. *Science* 297(5587):1708–1711. [PubMed: 12215649]
114. Barraclough DJ, Conroy ML, & Lee D (2004) Prefrontal cortex and decision making in a mixed-strategy game. *Nat Neurosci* 7(4):404–410. [PubMed: 15004564]
115. Kim JN & Shadlen MN (1999) Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nature Neuroscience* 2(2):176–185. [PubMed: 10195203]
116. Leon MI & Shadlen MN (1999) Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. *Neuron* 24(2):415–425. [PubMed: 10571234]
117. Averbeck BB, Chafee MV, Crowe DA, & Georgopoulos AP (2002) Parallel processing of serial movements in prefrontal cortex. *Proc Natl Acad Sci U S A* 99(20):13172–13177. [PubMed: 12242330]
118. Berdyeva TK & Olson CR (2010) Rank signals in four areas of macaque frontal cortex during selection of actions and objects in serial order. *J Neurophysiol* 104(1):141–159. [PubMed: 20445037]
119. Sigala N, Kusunoki M, Nimmo-Smith I, Gaffan D, & Duncan J (2008) Hierarchical coding for sequential task events in the monkey prefrontal cortex. *Proc Natl Acad Sci U S A* 105(33):11969–11974. [PubMed: 18689686]
120. Mante V, Sussillo D, Shenoy KV, & Newsome WT (2013) Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503(7474):78–84. [PubMed: 24201281]
121. Parthasarathy A, et al. (2017) Mixed selectivity morphs population codes in prefrontal cortex. *Nat Neurosci* 20(12):1770–1779. [PubMed: 29184197]
122. Dang W, Jaffe RJ, Qi XL, & Constantinidis C (2020) Emergence of Non-Linear Mixed Selectivity in Prefrontal Cortex after Training. *BioRxiv* 10.1101/2020.08.02.233247.
123. Lara AH & Wallis JD (2014) Executive control processes underlying multi-item working memory. *Nat Neurosci* 17(6):876–883.
124. Konecky RO, Smith MA, & Olson CR (2017) Monkey prefrontal neurons during Sternberg task performance: full contents of working memory or most recent item? *J Neurophysiol* 117(6):2269–2281. [PubMed: 28331006]
125. Lafer-Sousa R & Conway BR (2013) Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. *Nat Neurosci* 16(12):1870–1878. [PubMed: 24141314]

126. Tang H, Riley MR, & Constantinidis C (2017) Lateralization of Executive Function: Working Memory Advantage for Same Hemifield Stimuli in the Monkey. *Front Neurosci* 11:532. [PubMed: 29018321]
127. Lauwereyns J, et al. (2001) Responses to task-irrelevant visual features by primate prefrontal neurons. *J Neurophysiol* 86(4):2001–2010. [PubMed: 11600657]
128. Donahue CH & Lee D (2015) Dynamic routing of task-relevant signals for decision making in dorsolateral prefrontal cortex. *Nat Neurosci* 18(2):295–301. [PubMed: 25581364]
129. Buzsaki G (2010) Neural syntax: cell assemblies, synapsembles, and readers. *Neuron* 68(3):362–385. [PubMed: 21040841]
130. Roux F & Uhlhaas PJ (2014) Working memory and neural oscillations: alpha–gamma versus theta–gamma codes for distinct WM information? *Trends in Cognitive Sciences* 18(1):16–25. [PubMed: 24268290]
131. Liebe S, Hoerzer GM, Logothetis NK, & Rainer G (2012) Theta coupling between V4 and prefrontal cortex predicts visual short-term memory performance. *Nat Neurosci* 15(3):456–462, S451–452. [PubMed: 22286175]
132. Brincat SL & Miller EK (2015) Frequency-specific hippocampal-prefrontal interactions during associative learning. *Nat Neurosci* 18(4):576–581. [PubMed: 25706471]
133. Salazar RF, Dotson NM, Bressler SL, & Gray CM (2012) Content-specific fronto-parietal synchronization during visual working memory. *Science* 338(6110):1097–1100. [PubMed: 23118014]
134. Buschman TJ, Denovellis EL, Diogo C, Bullock D, & Miller EK (2012) Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. *Neuron* 76(4):838–846. [PubMed: 23177967]
135. Siegel M, Warden MR, & Miller EK (2009) Phase-dependent neuronal coding of objects in short-term memory. *Proc Natl Acad Sci U S A* 106(50):21341–21346. [PubMed: 19926847]
136. Lundqvist M, Herman P, Warden MR, Brincat SL, & Miller EK (2018) Gamma and beta bursts during working memory readout suggest roles in its volitional control. *Nat Commun* 9(1):394. [PubMed: 29374153]
137. Wutz A, Loonis R, Roy JE, Donoghue JA, & Miller EK (2018) Different Levels of Category Abstraction by Different Dynamics in Different Prefrontal Areas. *Neuron* 97(3):716–726 e718. [PubMed: 29395915]
138. Bastos AM, Loonis R, Kornblith S, Lundqvist M, & Miller EK (2018) Laminar recordings in frontal cortex suggest distinct layers for maintenance and control of working memory. *Proc Natl Acad Sci U S A* 115(5):1117–1122. [PubMed: 29339471]
139. Stanley DA, Roy JE, Aoi MC, Kopell NJ, & Miller EK (2018) Low-Beta Oscillations Turn Up the Gain During Category Judgments. *Cereb Cortex* 28(1):116–130. [PubMed: 29253255]
140. Miller EK, Lundqvist M, & Bastos AM (2018) Working Memory 2.0. *Neuron* 100(2):463–475. [PubMed: 30359609]
141. Michalareas G, et al. (2016) Alpha-Beta and Gamma Rhythms Subserve Feedback and Feedforward Influences among Human Visual Cortical Areas. *Neuron* 89(2):384–397. [PubMed: 26777277]
142. Bauer M, Oostenveld R, Peeters M, & Fries P (2006) Tactile spatial attention enhances gamma-band activity in somatosensory cortex and reduces low-frequency activity in parieto-occipital areas. *J Neurosci* 26(2):490–501. [PubMed: 16407546]
143. Womelsdorf T, Fries P, Mitra PP, & Desimone R (2006) Gamma-band synchronization in visual cortex predicts speed of change detection. *Nature* 439(7077):733–736. [PubMed: 16372022]
144. Desimone R & Duncan J (1995) Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18:193–222. [PubMed: 7605061]
145. Pesaran B, Pezaris JS, Sahani M, Mitra PP, & Andersen RA (2002) Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat Neurosci* 5(8):805–811. [PubMed: 12134152]
146. Warden MR & Miller EK (2007) The representation of multiple objects in prefrontal neuronal delay activity. *Cereb Cortex* 17 Suppl 1:i41–50. [PubMed: 17726003]
147. Stokes MG, et al. (2013) Dynamic coding for cognitive control in prefrontal cortex. *Neuron* 78(2):364–375. [PubMed: 23562541]

148. Crowe DA, Averbek BB, & Chafee MV (2008) Neural ensemble decoding reveals a correlate of viewer- to object-centered spatial transformation in monkey parietal cortex. *J Neurosci* 28(20):5218–5228. [PubMed: 18480278]
149. Spaak E, Watanabe K, Funahashi S, & Stokes MG (2017) Stable and Dynamic Coding for Working Memory in Primate Prefrontal Cortex. *J Neurosci* 37(27):6503–6516. [PubMed: 28559375]
150. Parthasarathy A, et al. (2019) Time-invariant working memory representations in the presence of code-morphing in the lateral prefrontal cortex. *Nat Commun* 10(1):4995. [PubMed: 31676790]
151. Grill-Spector K, Henson R, & Martin A (2006) Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci* 10(1):14–23. [PubMed: 16321563]
152. Miller EK, Li L, & Desimone R (1991) A neural mechanism for working and recognition memory in inferior temporal cortex. *Science* 254(5036):1377–1379. [PubMed: 1962197]
153. Steinmetz MA, Connor CE, Constantinidis C, & McLaughlin JR (1994) Covert attention suppresses neuronal responses in area 7a of the posterior parietal cortex. *Journal of Neurophysiology* 72(2):1020–1023. [PubMed: 7983507]
154. Woloszyn L & Sheinberg DL (2009) Neural dynamics in inferior temporal cortex during a visual working memory task. *J Neurosci* 29(17):5494–5507. [PubMed: 19403817]
155. Qi XL, Meyer T, Stanford TR, & Constantinidis C (2012) Neural correlates of a decision variable before learning to perform a Match/Nonmatch task. *Journal of Neuroscience* 32(18):6161–6169. [PubMed: 22553022]
156. Sugase-Miyamoto Y, Liu Z, Wiener MC, Optican LM, & Richmond BJ (2008) Short-term memory trace in rapidly adapting synapses of inferior temporal cortex. *PLoS Comput Biol* 4(5):e1000073. [PubMed: 18464917]
157. Sinz FH, Pitkow X, Reimer J, Bethge M, & Tolias AS (2019) Engineering a Less Artificial Intelligence. *Neuron* 103(6):967–979. [PubMed: 31557461]
158. LeCun Y, Bengio Y, & Hinton G (2015) Deep learning. *Nature* 521(7553):436–444. [PubMed: 26017442]
159. Yang GR, Joglekar MR, Song HF, Newsome WT, & Wang XJ (2019) Task representations in neural networks trained to perform many cognitive tasks. *Nat Neurosci* 22(2):297–306. [PubMed: 30643294]
160. Masse NY, Yang GR, Song HF, Wang XJ, & Freedman DJ (2019) Circuit mechanisms for the maintenance and manipulation of information in working memory. *Nat Neurosci* 22(7):1159–1167. [PubMed: 31182866]
161. Kim R & Sejnowski TJ (2021) Strong inhibitory signaling underlies stable temporal dynamics and working memory in spiking neural networks. *Nat Neurosci* 24(1):129–139. [PubMed: 33288909]
162. Fiebig F & Lansner A (2017) A Spiking Working Memory Model Based on Hebbian Short-Term Potentiation. *J Neurosci* 37(1):83–96. [PubMed: 28053032]
163. Hansel D & Mato G (2013) Short-term plasticity explains irregular persistent activity in working memory tasks. *J Neurosci* 33(1):133–149. [PubMed: 23283328]
164. Barbosa J, et al. (2020) Interplay between persistent activity and activity-silent dynamics in the prefrontal cortex underlies serial biases in working memory. *Nat Neurosci* 23(8):1016–1024. [PubMed: 32572236]
165. Stein H, et al. (2020) Reduced serial dependence suggests deficits in synaptic potentiation in anti-NMDAR encephalitis and schizophrenia. *Nat Commun* 11(1):4250. [PubMed: 32843635]
166. Fischer J & Whitney D (2014) Serial dependence in visual perception. *Nat Neurosci* 17(5):738–743. [PubMed: 24686785]
167. Papadimitriou C, Ferdoash A, & Snyder LH (2015) Ghosts in the machine: memory interference from the previous trial. *J Neurophysiol* 113(2):567–577. [PubMed: 25376781]
168. Bliss DP, Sun JJ, & D'Esposito M (2017) Serial dependence is absent at the time of perception but increases in visual working memory. *Sci Rep* 7(1):14739. [PubMed: 29116132]
169. Bliss DP & D'Esposito M (2017) Synaptic augmentation in a cortical circuit model reproduces serial dependence in visual working memory. *PLoS One* 12(12):e0188927. [PubMed: 29244810]

170. Kilpatrick ZP (2018) Synaptic mechanisms of interference in working memory. *Sci Rep* 8(1):7879. [PubMed: 29777113]
171. Elston GN (2003) The pyramidal neuron in occipital, temporal and prefrontal cortex of the owl monkey (*Aotus trivirgatus*): regional specialization in cell structure. *Eur J Neurosci* 17(6):1313–1318. [PubMed: 12670321]
172. Elston GN (2000) Pyramidal cells of the frontal lobe: all the more spinous to think with. *J Neurosci* 20(18):RC95. [PubMed: 10974092]
173. Katsuki F, et al. (2014) Differences in intrinsic functional organization between dorsolateral prefrontal and posterior parietal cortex. *Cereb Cortex* 24(9):2334–2349. [PubMed: 23547137]
174. Hart E & Huk AC (2020) Recurrent circuit dynamics underlie persistent activity in the macaque frontoparietal network. *Elife* 9.
175. Glasser MF & Van Essen DC (2011) Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. *J Neurosci* 31(32):11597–11616. [PubMed: 21832190]
176. Huntenburg JM, et al. (2017) A Systematic Relationship Between Functional Connectivity and Intracortical Myelin in the Human Cerebral Cortex. *Cereb Cortex* 27(2):981–997. [PubMed: 28184415]
177. Burt JB, et al. (2018) Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nat Neurosci* 21(9):1251–1259. [PubMed: 30082915]
178. Goldman-Rakic PS (1984) Modular organization of prefrontal cortex. *Trends in Neurosciences* 7(11):419–424.
179. Levitt JB, Lewis DA, Yoshioka T, & Lund JS (1993) Topography of pyramidal neuron intrinsic connections in macaque monkey prefrontal cortex (areas 9 and 46). *Journal of Comparative Neurology* 338(3):360–376.
180. Kritzer MF & Goldman-Rakic PS (1995) Intrinsic circuit organization of the major layers and sublayers of the dorsolateral prefrontal cortex in the rhesus monkey. *Journal of Comparative Neurology* 359(1):131–143.
181. Pucak ML, Levitt JB, Lund JS, & Lewis DA (1996) Patterns of intrinsic and associational circuitry in monkey prefrontal cortex. *Journal of Comparative Neurology* 376(4):614–630.
182. Constantinidis C & Wang XJ (2004) A neural circuit basis for spatial working memory. *Neuroscientist* 10(6):553–565. [PubMed: 15534040]
183. Wang XJ (2001) Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci* 24(8):455–463. [PubMed: 11476885]
184. Wang XJ (1999) Synaptic Basis of Cortical Persistent Activity: the Importance of NMDA Receptors to Working Memory. *J. Neurosci* 19(21):9587–9603. [PubMed: 10531461]
185. Zick JL, et al. (2018) Blocking NMDAR Disrupts Spike Timing and Decouples Monkey Prefrontal Circuits: Implications for Activity-Dependent Disconnection in Schizophrenia. *Neuron* 98(6):1243–1255 e1245. [PubMed: 29861281]
186. Wang M, et al. (2013) NMDA receptors subserve persistent neuronal firing during working memory in dorsolateral prefrontal cortex. *Neuron* 77(4):736–749. [PubMed: 23439125]
187. Wang M & Arnsten AF (2015) Contribution of NMDA receptors to dorsolateral prefrontal cortical networks in primates. *Neurosci Bull* 31(2):191–197. [PubMed: 25754145]
188. Rodermund P, Westendorff S, & Nieder A (2020) Blockage of NMDA- and GABA(A) Receptors Improves Working Memory Selectivity of Primate Prefrontal Neurons. *J Neurosci* 40(7):1527–1537. [PubMed: 31911457]
189. Wang M, Wong AH, & Liu F (2012) Interactions between NMDA and dopamine receptors: a potential therapeutic target. *Brain Res* 1476:154–163. [PubMed: 22472597]
190. Levitt P, Rakic P, & Goldman-Rakic P (1984) Region-specific distribution of catecholamine afferents in primate cerebral cortex: a fluorescence histochemical analysis. *Journal of Comparative Neurology* 227(1):23–36.
191. Constantinidis C & Goldman-Rakic PS (2002) Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. *Journal of Neurophysiology* 88:3487–3497. [PubMed: 12466463]

192. Rao SG, Williams GV, & Goldman-Rakic PS (2000) Destruction and creation of spatial tuning by disinhibition: GABA(A) blockade of prefrontal cortical neurons engaged by working memory. *Journal of Neuroscience* 20(1):485–494. [PubMed: 10627624]
193. Constantinidis C, Williams GV, & Goldman-Rakic PS (2002) A role for inhibition in shaping the temporal flow of information in prefrontal cortex. *Nature Neuroscience* 5(2):175–180. [PubMed: 11802172]
194. Compte A, Brunel N, Goldman-Rakic PS, & Wang XJ (2000) Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex* 10(9):910–923. [PubMed: 10982751]
195. Wang XJ, Tegner J, Constantinidis C, & Goldman-Rakic PS (2004) Division of labor among distinct subtypes of inhibitory neurons in a cortical microcircuit of working memory. *Proc Natl Acad Sci U S A* 101(5):1368–1373. [PubMed: 14742867]
196. Gabbott PL & Bacon SJ (1997) Vasoactive intestinal polypeptide containing neurones in monkey medial prefrontal cortex (mPFC): colocalisation with calretinin. *Brain Res* 744(1):179–184. [PubMed: 9030431]
197. Li S, Zhou X, Constantinidis C, & Qi XL (2020) Plasticity of Persistent Activity and Its Constraints. *Front Neural Circuits* 14:15. [PubMed: 32528254]
198. Meskenaitė V (1997) Calretinin-immunoreactive local circuit neurons in area 17 of the cynomolgus monkey, *Macaca fascicularis*. *J Comp Neurol* 379(1):113–132. [PubMed: 9057116]
199. Fish KN, Rocco BR, & Lewis DA (2018) Laminar Distribution of Subsets of GABAergic Axon Terminals in Human Prefrontal Cortex. *Front Neuroanat* 12:9. [PubMed: 29503610]
200. Melchitzky DS & Lewis DA (2008) Dendritic-targeting GABA neurons in monkey prefrontal cortex: comparison of somatostatin- and calretinin-immunoreactive axon terminals. *Synapse* 62(6):456–465. [PubMed: 18361442]
201. Defelipe J, Gonzalez-Albo MC, Del Rio MR, & Elston GN (1999) Distribution and patterns of connectivity of interneurons containing calbindin, calretinin, and parvalbumin in visual areas of the occipital and temporal lobes of the macaque monkey. *J Comp Neurol* 412(3):515–526. [PubMed: 10441237]
202. Elston GN & Gonzalez-Albo MC (2003) Parvalbumin-, calbindin-, and calretinin-immunoreactive neurons in the prefrontal cortex of the owl monkey (*Aotus trivirgatus*): a standardized quantitative comparison with sensory and motor areas. *Brain Behav Evol* 62(1):19–30. [PubMed: 12907857]
203. Torres-Gomez S, et al. (2020) Changes in the Proportion of Inhibitory Interneuron Types from Sensory to Executive Areas of the Primate Neocortex: Implications for the Origins of Working Memory Representations. *Cereb Cortex* 30(8):4544–4562. [PubMed: 32227119]
204. Zhou X, Katsuki F, Qi XL, & Constantinidis C (2012) Neurons with inverted tuning during the delay periods of working memory tasks in the dorsal prefrontal and posterior parietal cortex. *J Neurophysiol* 108(1):31–38. [PubMed: 22490554]
205. Sawaguchi T & Goldman-Rakic PS (1994) The role of D1-dopamine receptor in working memory: local injections of dopamine antagonists into the prefrontal cortex of rhesus monkeys performing an oculomotor delay ed-response task. *Journal of Neurophysiology* 71(2):515–528. [PubMed: 7909839]
206. Williams GV & Goldman-Rakic PS (1995) Modulation of memory fields by dopamine D1 receptors in prefrontal cortex. *Nature* 376(6541):572–575. [PubMed: 7637804]
207. Vijayraghavan S, Wang M, Birnbaum SG, Williams GV, & Arnsten AF (2007) Inverted-U dopamine D1 receptor actions on prefrontal neurons engaged in working memory. *Nat Neurosci* 10(3):376–384. [PubMed: 17277774]
208. Ott T, Jacob SN, & Nieder A (2014) Dopamine receptors differentially enhance rule coding in primate prefrontal cortex neurons. *Neuron* 84(6):1317–1328. [PubMed: 25482027]
209. Jacob SN, Stalter M, & Nieder A (2016) Cell-type-specific modulation of targets and distractors by dopamine D1 receptors in primate prefrontal cortex. *Nat Commun* 7:13218. [PubMed: 27807366]
210. Wang M, Vijayraghavan S, & Goldman-Rakic PS (2004) Selective D2 receptor actions on the functional circuitry of working memory. *Science* 303(5659):853–856. [PubMed: 14764884]

211. Yang CR & Seamans JK (1996) Dopamine D1 receptor actions in layers V-VI rat prefrontal cortex neurons in vitro: modulation of dendritic-somatic signal integration. *J Neurosci* 16(5):1922–1935. [PubMed: 8774459]
212. Durstewitz D, Seamans JK, & Sejnowski TJ (2000) Neurocomputational models of working memory. *Nat Neurosci* 3(Suppl):1184–1191. [PubMed: 11127836]
213. Seamans JK, Durstewitz D, Christie BR, Stevens CF, & Sejnowski TJ (2001) Dopamine D1/D5 receptor modulation of excitatory synaptic inputs to layer V prefrontal cortex neurons. *Proc Natl Acad Sci U S A* 98(1):301–306. [PubMed: 11134516]
214. Chen G, Greengard P, & Yan Z (2004) Potentiation of NMDA receptor currents by dopamine D1 receptors in prefrontal cortex. *Proc Natl Acad Sci U S A* 101(8):2596–2600. [PubMed: 14983054]
215. Wang M, et al. (2007) Alpha2A-adrenoceptors strengthen working memory networks by inhibiting cAMP-HCN channel signaling in prefrontal cortex. *Cell* 129(2):397–410. [PubMed: 17448997]
216. Bartus RT (2000) On neurodegenerative diseases, models, and treatment strategies: lessons learned and lessons forgotten a generation following the cholinergic hypothesis. *Exp Neurol* 163(2):495–529. [PubMed: 10833325]
217. Zhou X, et al. (2011) Cholinergic modulation of working memory activity in primate prefrontal cortex. *J Neurophysiol* 106(5):2180–2188. [PubMed: 21795623]
218. Yang Y, et al. (2013) Nicotinic alpha7 receptors enhance NMDA cognitive circuits in dorsolateral prefrontal cortex. *Proc Natl Acad Sci U S A* 110(29):12078–12083. [PubMed: 23818597]
219. Liu R, et al. (2017) Intermittent Stimulation of the Nucleus Basalis of Meynert Improves Working Memory in Adult Monkeys. *Curr Biol* 7(17):2640–2646.
220. Qi XL, Liu R, Vazdarjanova AI, Blake DT, & Constantinidis C (2019) Nucleus Basalis Stimulation Enhances Working Memory and Stabilizes Attractor Networks in Prefrontal Cortex. *bioRxiv*:674465.
221. Constantinidis C & Procyk E (2004) The primate working memory networks. *Cogn Affect Behav Neurosci* 4(4):444–465. [PubMed: 15849890]
222. Pasternak T & Greenlee MW (2005) Working memory in primate sensory systems. *Nat Rev Neurosci* 6(2):97–107. [PubMed: 15654324]
223. Harrison SA & Tong F (2009) Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458(7238):632–635. [PubMed: 19225460]
224. Rawley JB & Constantinidis C (2009) Neural correlates of learning and working memory in the primate posterior parietal cortex. *Neurobiol Learn Mem* 91:129–138. [PubMed: 19116173]
225. Jonides J, et al. (1993) Spatial working memory in humans as revealed by PET. *Nature* 363(6430):623–625. [PubMed: 8510752]
226. Courtney SM, Ungerleider LG, Keil K, & Haxby JV (1997) Transient and sustained activity in a distributed neural system for human working memory. *Nature* 386(6625):608–611. [PubMed: 9121584]
227. Ungerleider LG, Courtney SM, & Haxby JV (1998) A neural system for human visual working memory. *Proceedings of the National Academy of Sciences of the United States of America* 95(3):883–890. [PubMed: 9448255]
228. Marshuetz C, Smith EE, Jonides J, DeGutis J, & Chenevert TL (2000) Order information in working memory: fMRI evidence for parietal and prefrontal mechanisms. *J Cogn Neurosci* 12(Suppl 2):130–144. [PubMed: 11506653]
229. Bunge SA, Ochsner KN, Desmond JE, Glover GH, & Gabrieli JD (2001) Prefrontal regions involved in keeping information in and out of mind. *Brain* 124(Pt 10):2074–2086. [PubMed: 11571223]
230. Owen AM, et al. (1998) Functional organization of spatial and nonspatial working memory processing within the human lateral frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America* 95(13):7721–7726. [PubMed: 9636217]
231. Stem CE, Sherman SJ, Kirchoff BA, & Hasselmo ME (2001) Medial temporal and prefrontal contributions to working memory tasks with novel and familiar stimuli. *Hippocampus* 11(4):337–346. [PubMed: 11530838]

232. Chafee MV & Goldman-Rakic PS (1998) Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *Journal of Neurophysiology* 79(6):2919–2940. [PubMed: 9636098]
233. Constantinidis C & Steinmetz MA (1996) Neuronal activity in posterior parietal area 7a during the delay periods of a spatial memory task. *Journal of Neurophysiology* 76(2):1352–1355. [PubMed: 8871242]
234. Gnadt JW & Andersen RA (1988) Memory related motor planning activity in posterior parietal cortex of macaque. *Experimental Brain Research* 70(1):216–220. [PubMed: 3402565]
235. Fuster JM & Jervey JP (1982) Neuronal firing in the inferotemporal cortex of the monkey in a visual memory task. *Journal of Neuroscience* 2(3):361–375. [PubMed: 7062115]
236. Fuster JM & Jervey JP (1981) Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science* 212(4497):952–955. [PubMed: 7233192]
237. Miller EK, Li L, & Desimone R (1993) Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *Journal of Neuroscience* 13(4):1460–1478. [PubMed: 8463829]
238. Miyashita Y & Chang HS (1988) Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature* 331(6151):68–70. [PubMed: 3340148]
239. Naya Y, Yoshida M, & Miyashita Y (2001) Backward spreading of memory-retrieval signal in the primate temporal cortex. *Science* 291(5504):661–664. [PubMed: 11158679]
240. Nakamura K & Kubota K (1995) Mnemonic firing of neurons in the monkey temporal pole during a visual recognition memory task. *J Neurophysiol* 74(1):162–178. [PubMed: 7472321]
241. Sigala N & Logothetis NK (2002) Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* 415(6869):318–320. [PubMed: 11797008]
242. Miller EK & Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167–202. [PubMed: 11283309]
243. Cowan N (1988) Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychol Bull.* 104(2):163–191. [PubMed: 3054993]
244. D'Esposito M (2007) From cognitive to neural models of working memory. *Philos Trans R Soc Lond B Biol Sci.* 362(1481):761–772. [PubMed: 17400538]
245. Postle BR (2006) Working memory as an emergent property of the mind and brain. *Neuroscience* 139(1):23–38. [PubMed: 16324795]
246. Hazy TE, Frank MJ, O'Reilly RC (2006) Banishing the homunculus: making working memory work. *Neuroscience* 139(1):105–118. [PubMed: 16343792]
247. di Pellegrino G & Wise SP (1993) Effects of attention on visuomotor activity in the premotor and prefrontal cortex of a primate. *Somatosensory & Motor Research* 10(3):245–262. [PubMed: 8237213]
248. Suzuki M & Gottlieb J (2013) Distinct neural mechanisms of distractor suppression in the frontal and parietal lobe. *Nat Neurosci* 16(1):98–104. [PubMed: 23242309]
249. Qi XL, Elworthy AC, Lambert BC, & Constantinidis C (2015) Representation of remembered stimuli and task information in the monkey dorsolateral prefrontal and posterior parietal cortex. *J Neurophysiol* 113(1):44–57. [PubMed: 25298389]
250. Jacob SN & Nieder A (2014) Complementary roles for primate frontal and parietal cortex in guarding working memory from distractor stimuli. *Neuron* 83(1):226–237. [PubMed: 24991963]
251. Ibos G, Duhamel JR, & Ben Hamed S (2013) A functional hierarchy within the parietofrontal network in stimulus selection and attention control. *J Neurosci* 33(19):8359–8369. [PubMed: 23658175]
252. Swaminathan SK & Freedman DJ (2012) Preferential encoding of visual categories in parietal cortex compared with prefrontal cortex. *Nat Neurosci* 15(2):315–320. [PubMed: 22246435]
253. Crowe DA, et al. (2013) Prefrontal neurons transmit signals to parietal neurons that reflect executive control of cognition. *Nat Neurosci* 16(10):1484–1491. [PubMed: 23995071]

254. Katsuki F & Constantinidis C (2012) Unique and shared roles of the posterior parietal and dorsolateral prefrontal cortex in cognitive functions. *Frontiers in Integrative Neuroscience* 6:17. [PubMed: 22563310]
255. Chafee MV & Goldman-Rakic PS (2000) Inactivation of parietal and prefrontal cortex reveals interdependence of neural activity during memory-guided saccades. *J Neurophysiol* 83(3):1550–1566. [PubMed: 10712479]
256. Sommer MA & Tehovnik EJ (1997) Reversible inactivation of macaque frontal eye field. *Experimental Brain Research* 116(2):229–249. [PubMed: 9348123]
257. Dias EC & Segraves MA (1999) Muscimol-induced inactivation of monkey frontal eye field: effects on visually and memory-guided saccades. *Journal of Neurophysiology* 81(5):2191–2214. [PubMed: 10322059]
258. Wilke M, Kagan I, & Andersen RA (2012) Functional imaging reveals rapid reorganization of cortical activity after parietal inactivation in monkeys. *Proc Natl Acad Sci U S A* 109(21):8274–8279. [PubMed: 22562793]
259. Li CS, Mazzoni P, & Andersen RA (1999) Effect of reversible inactivation of macaque lateral intraparietal area on visual and memory saccades. *Journal of Neurophysiology* 81(4):1827–1838. [PubMed: 10200217]
260. Wardak C, Olivier E, & Duhamel JR (2002) Saccadic target selection deficits after lateral intraparietal area inactivation in monkeys. *J Neurosci* 22(22):9877–9884. [PubMed: 12427844]
261. Wardak C, Olivier E, & Duhamel JR (2004) A deficit in covert attention after parietal cortex inactivation in the monkey. *Neuron* 42(3):501–508. [PubMed: 15134645]
262. Liu Y, Yttri EA, & Snyder LH (2010) Intention and attention: different functional roles for LIPd and LIPv. *Nat Neurosci* 13(4):495–500. [PubMed: 20190746]
263. Funahashi S, Bruce CJ, & Goldman-Rakic PS (1993) Dorsolateral prefrontal lesions and oculomotor delayed-response performance: evidence for mnemonic "scotomas". *Journal of Neuroscience* 13(4):1479–1497. [PubMed: 8463830]
264. Funahashi S (2015) Functions of delay-period activity in the prefrontal cortex and mnemonic scotomas revisited. *Front Syst Neurosci* 9:2. [PubMed: 25698942]
265. Offen S, Schluppeck D, & Heeger DJ (2009) The role of early visual cortex in visual short-term memory and visual attention. *Vision research* 49(10):1352–1362. [PubMed: 18329065]
266. Xing Y, Ledgeway T, McGraw PV, & Schluppeck D (2013) Decoding working memory of stimulus contrast in early visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33(25):10301–10311. [PubMed: 23785144]
267. Albers AM, Kok P, Toni I, Dijkerman HC, & de Lange FP (2013) Shared representations for working memory and mental imagery in early visual cortex. *Current biology : CB* 23(15):1427–1431. [PubMed: 23871239]
268. Sreenivasan KK, Vytlačil J, & D'Esposito M (2014) Distributed and Dynamic Storage of Working Memory Stimulus Information in Extrastriate Cortex. *Journal of Cognitive Neuroscience* 26(5):1141–1153. [PubMed: 24392897]
269. Ester EF, Anderson DE, Serences JT, & Awh E (2013) A Neural Measure of Precision in Visual Working Memory. *Journal of Cognitive Neuroscience* 25(5):754–761. [PubMed: 23469889]
270. Tong F & Pratte MS (2012) Decoding patterns of human brain activity. *Annual review of psychology* 63:483–509.
271. Bergmann J, Genc E, Kohler A, Singer W, & Pearson J (2014) Neural Anatomy of Primary Visual Cortex Limits Visual Working Memory. *Cerebral cortex*.
272. Ester EF, Sprague TC, & Serences JT (2015) Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. *Neuron* 87(4):893–905. [PubMed: 26257053]
273. Christophel TB, Klink PC, Spitzer B, Roelfsema PR, & Haynes JD (2017) The Distributed Nature of Working Memory. *Trends Cogn Sci* 21(2):111–124. [PubMed: 28063661]
274. Super H, Spekreijse H, & Lamme VA (2001) A neural correlate of working memory in the monkey primary visual cortex. *Science* 293(5527):120–124. [PubMed: 11441187]
275. Roelfsema PR (2015) The role of the different layers of primary visual cortex in working memory. *Journal of vision* 15(12):1406.

276. Logothetis NK & Wandell BA (2004) Interpreting the BOLD signal. *Annu Rev Physiol* 66:735–769. [PubMed: 14977420]
277. Alexander GE, DeLong MR, & Strick PL (1986) Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu Rev Neurosci* 9:357–381. [PubMed: 3085570]
278. Middleton FA & Strick PL (2002) Basal-ganglia 'projections' to the prefrontal cortex of the primate. *Cereb Cortex* 12(9):926–935. [PubMed: 12183392]
279. Haber SN, Kunishio K, Mizobuchi M, & Lynd-Balta E (1995) The orbital and medial prefrontal circuit through the primate basal ganglia. *J Neurosci* 15(7 Pt 1):4851–4867. [PubMed: 7623116]
280. Selemon LD & Goldman-Rakic PS (1985) Longitudinal topography and interdigitation of corticostriatal projections in the rhesus monkey. *J Neurosci* 5(3):776–794. [PubMed: 2983048]
281. Hadj-Bouziane F, Meunier M, & Boussaoud D (2003) Conditional visuo-motor learning in primates: a key role for the basal ganglia. *J Physiol Paris* 97(4-6):567–579. [PubMed: 15242666]
282. Hikosaka O, Sakamoto M, & Usui S (1989) Functional properties of monkey caudate neurons. I. Activities related to saccadic eye movements. *J Neurophysiol* 61(4):780–798. [PubMed: 2723720]
283. Hikosaka O, Sakamoto M, & Usui S (1989) Functional properties of monkey caudate neurons. III. Activities related to expectation of target and reward. *J Neurophysiol* 61(4):814–832. [PubMed: 2723722]
284. Kawagoe R, Takikawa Y, & Hikosaka O (1998) Expectation of reward modulates cognitive signals in the basal ganglia. *Nat Neurosci* 1(5):411–416. [PubMed: 10196532]
285. Kermadi I & Joseph JP (1995) Activity in the caudate nucleus of monkey during spatial sequencing. *J Neurophysiol* 74(3):911–933. [PubMed: 7500161]
286. Wise SP, Murray EA, & Gerfen CR (1996) The frontal cortex-basal ganglia system in primates. *Critical Reviews in Neurobiology* 10(3&4):317–356. [PubMed: 8978985]
287. Llinas R, Ribary U, Contreras D, & Pedroarena C (1998) The neuronal basis for consciousness. *Philos Trans R Soc Lond B Biol Sci* 353(1377):1841–1849. [PubMed: 9854256]
288. Llinas RR, Leznik E, & Urbano FJ (2002) Temporal binding via cortical coincidence detection of specific and nonspecific thalamocortical inputs: a voltage-dependent dye-imaging study in mouse brain slices. *Proc Natl Acad Sci U S A* 99(1):449–454. [PubMed: 11773628]
289. Houk JC & Wise SP (1995) Distributed modular architectures linking basal ganglia, cerebellum, and cerebral cortex: their role in planning and controlling action. *Cereb Cortex* 5(2):95–110. [PubMed: 7620294]
290. Romanski LM, Giguere M, Bates JF, & Goldman-Rakic PS (1997) Topographic organization of medial pulvinar connections with the prefrontal cortex in the rhesus monkey. *J Comp Neurol* 379(3):313–332. [PubMed: 9067827]
291. Selemon LD & Goldman-Rakic PS (1988) Common cortical and subcortical targets of the dorsolateral prefrontal and posterior parietal cortices in the rhesus monkey: evidence for a distributed neural network subserving spatially guided behavior. *J Neurosci* 8(11):4049–4068. [PubMed: 2846794]
292. Giguere M & Goldman-Rakic PS (1988) Mediodorsal nucleus: areal, laminar, and tangential distribution of afferents and efferents in the frontal lobe of rhesus monkeys. *J Comp Neurol* 277:195–213. [PubMed: 2466057]
293. Goldman-Rakic PS & Porrino LJ (1985) The primate mediodorsal (MD) nucleus and its projection to the frontal lobe. *J Comp Neurol* 242(4):535–560. [PubMed: 2418080]
294. Rouiller EM, Tanne J, Moret V, & Boussaoud D (1999) Origin of thalamic inputs to the primary, premotor, and supplementary motor cortical areas and to area 46 in macaque monkeys: a multiple retrograde tracing study. *J Comp Neurol* 409(1):131–152. [PubMed: 10363716]
295. Tanibuchi I & Goldman-Rakic PS (2003) Dissociation of spatial-, object-, and sound-coding neurons in the mediodorsal nucleus of the primate thalamus. *J Neurophysiol* 89(2):1067–1077. [PubMed: 12574481]
296. Fuster JM & Alexander GE (1973) Firing changes in cells of the nucleus medialis dorsalis associated with delayed response behavior. *Brain Res* 61:79–91. [PubMed: 4204130]

297. Watanabe Y & Funahashi S (2004) Neuronal activity throughout the primate mediodorsal nucleus of the thalamus during oculomotor delayed-responses. I. Cue-, delay-, and response-period activity. *J Neurophysiol* 92(3):1738–1755. [PubMed: 15140911]
298. Watanabe Y & Funahashi S (2004) Neuronal Activity Throughout the Primate Mediodorsal Nucleus of the Thalamus During Oculomotor Delayed-Responses. II. Activity Encoding Visual Versus Motor Signal. *J Neurophysiol* 92(3):1756–1769. [PubMed: 15140912]
299. Hampson RE, Pons TP, Stanford TR, & Deadwyler SA (2004) Categorization in the monkey hippocampus: a possible mechanism for encoding information into memory. *Proc Natl Acad Sci U S A* 101(9):3184–3189. Epub 2004 Feb 31. [PubMed: 14978264]
300. Mejias JF & Wang XJ (2019) Mechanisms of distributed working memory in a large-scale model of the macaque neocortex. *BioRxiv*:760231.
301. Walker AE (1940) A cytoarchitectural study of the prefrontal area of the macaque monkey. *Journal of Comparative Neurology* 73:59–86.
302. Preuss TM & Goldman-Rakic PS (1991) Architectonics of the parietal and temporal association cortex in the strepsirhine primate Galago compared to the anthropoid primate Macaca. *Journal of Comparative Neurology* 310(4):475–506.
303. Petrides M (2000) The role of the mid-dorsolateral prefrontal cortex in working memory. *Exp Brain Res* 133(1):44–54. [PubMed: 10933209]
304. Goulas A, et al. (2017) Intrinsic functional architecture of the macaque dorsal and ventral lateral frontal cortex. *J Neurophysiol* 117(3):1084–1099. [PubMed: 28003408]
305. Selemon LD & Goldman-Rakic PS (1988) Common cortical and subcortical targets of the dorsolateral prefrontal and posterior parietal cortices in the rhesus monkey: evidence for a distributed neural network subserving spatially guided behavior. *Journal of Neuroscience* 8(11):4049–4068. [PubMed: 2846794]
306. Cavada C & Goldman-Rakic PS (1989) Posterior parietal cortex in rhesus monkey: II. Evidence for segregated corticocortical networks linking sensory and limbic areas with the frontal lobe. *Journal of Comparative Neurology* 287(4):422–445.
307. Petrides M & Pandya DN (1984) Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. *Journal of Comparative Neurology* 228(1):105–116.
308. Borra E, et al. (2017) Rostro-caudal Connectional Heterogeneity of the Dorsal Part of the Macaque Prefrontal Area 46. *Cereb Cortex*.
309. Barbas H (2015) General cortical and special prefrontal connections: principles from structure to function. *Annu Rev Neurosci* 38:269–289. [PubMed: 25897871]
310. Gerbella M, Borra E, Tonelli S, Rozzi S, & Luppino G (2013) Connectional heterogeneity of the ventral part of the macaque area 46. *Cereb Cortex* 23(4):967–987. [PubMed: 22499799]
311. Ungerleider LG & Mishkin M (1982) Two cortical visual systems. *Analysis of Visual Behavior*, eds Ingle DJ, Goodale MA, & Mansfield RJW (MIT Press, Cambridge, MA), pp 549–586.
312. Felleman DJ & Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* 1(1):1–47. [PubMed: 1822724]
313. Rainer G, Asaad WF, & Miller EK (1998) Memory fields of neurons in the primate prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America* 95(25):15008–15013. [PubMed: 9844006]
314. Riley MR, Qi XL, & Constantinidis C (2017) Functional specialization of areas along the anterior-posterior axis of the primate prefrontal cortex. *Cereb Cortex* 27:3683–3697. [PubMed: 27371761]
315. Kadohisa M, et al. (2015) Spatial and temporal distribution of visual information coding in lateral prefrontal cortex. *Eur J Neurosci* 41(1):89–96. [PubMed: 25307044]
316. Noudoost B, Clark KL, & Moore T (2014) A distinct contribution of the frontal eye field to the visual representation of saccadic targets. *J Neurosci* 34(10):3687–3698. [PubMed: 24599467]
317. Sawaguchi T & Iba M (2001) Prefrontal cortical representation of visuospatial working memory in monkeys examined by local inactivation with muscimol. *J Neurophysiol* 86(4):2041–2053. [PubMed: 11600660]
318. Bichot NP, Heard MT, DeGennaro EM, & Desimone R (2015) A Source for Feature-Based Attention in the Prefrontal Cortex. *Neuron* 88(4):832–844. [PubMed: 26526392]

319. Arnsten AF (2013) The neurobiology of thought: the groundbreaking discoveries of Patricia Goldman-Rakic 1937-2003. *Cereb Cortex* 23(10):2269–2281. [PubMed: 23926115]
320. Clark KL, Noudoost B, & Moore T (2012) Persistent spatial information in the frontal eye field during object-based short-term memory. *J Neurosci* 32(32):10907–10914. [PubMed: 22875925]
321. Peng X, Sereno ME, Silva AK, Lehky SR, & Sereno AB (2008) Shape selectivity in primate frontal eye field. *J Neurophysiol* 100(2):796–814. [PubMed: 18497359]
322. Sereno AB & Maunsell JH (1998) Shape selectivity in primate lateral intraparietal cortex. *Nature* 395(6701):500–503. [PubMed: 9774105]
323. Janssen P, Srivastava S, Omelet S, & Orban GA (2008) Coding of shape and position in macaque lateral intraparietal area. *J Neurosci* 28(26):6679–6690. [PubMed: 18579742]
324. Gross CG, Rocha-Miranda CE, & Bender DB (1972) Visual properties of neurons in inferotemporal cortex of the Macaque. *J Neurophysiol* 35(1):96–111. [PubMed: 4621506]
325. Desimone R, Albright TD, Gross CG, & Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4(8):2051–2062. [PubMed: 6470767]
326. Tanaka K, Saito H, Fukada Y, & Moriya M (1991) Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J Neurophysiol* 66(1):170–189. [PubMed: 1919665]
327. Fujita I, Tanaka K, Ito M, & Cheng K (1992) Columns for visual features of objects in monkey inferotemporal cortex. *Nature* 360(6402):343–346. [PubMed: 1448150]
328. Constantinidis C & Steinmetz MA (2001) Neuronal responses in area 7a to multiple stimulus displays: I. Neurons encode the location of the salient stimulus. *Cerebral Cortex* 11:581–591. [PubMed: 11415960]
329. Tsao DY, Freiwald WA, Tootell RB, & Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. *Science* 311(5761):670–674. [PubMed: 16456083]
330. Chang L, Bao P, & Tsao DY (2017) The representation of colored objects in macaque color patches. *Nat Commun* 8(1):2064. [PubMed: 29234028]
331. Popivanov ID, Jastorff J, Vanduffel W, & Vogels R (2014) Heterogeneous single-unit selectivity in an fMRI-defined body-selective patch. *J Neurosci* 34(1):95–111. [PubMed: 24381271]
332. Puig MV & Miller EK (2012) The role of prefrontal dopamine D1 receptors in the neural mechanisms of associative learning. *Neuron* 74(5):874–886. [PubMed: 22681691]
333. Clark KL, Noudoost B, & Moore T (2014) Persistent spatial information in the FEF during object-based short-term memory does not contribute to task performance. *J Cogn Neurosci* 26(6):1292–1299. [PubMed: 24673408]
334. Rushworth MF, Nixon PD, Eacott MJ, & Passingham RE (1997) Ventral prefrontal cortex is not essential for working memory. *J Neurosci* 17(12):4829–4838. [PubMed: 9169541]
335. Petrides M (2005) Lateral prefrontal cortex: architectonic and functional organization. *Philos Trans R Soc Lond B Biol Sci* 360(1456):781–795. [PubMed: 15937012]
336. Petrides M (1996) Specialized systems for the processing of mnemonic information within the primate frontal cortex. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences* 351(1346):1455–1461; discussion 1461-1452. [PubMed: 8941957]
337. Buckley MJ, et al. (2009) Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. *Science* 325(5936):52–58. [PubMed: 19574382]
338. Rygula R, Walker SC, Clarke HF, Robbins TW, & Roberts AC (2010) Differential contributions of the primate ventrolateral prefrontal and orbitofrontal cortex to serial reversal learning. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30(43):14552–14559. [PubMed: 20980613]
339. Romo R, Brody CD, Hernandez A, & Lemus L (1999) Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* 399(6735):470–473. [PubMed: 10365959]
340. Wang L, et al. (2015) Differential roles of delay-period neural activity in the monkey dorsolateral prefrontal cortex in visual-haptic crossmodal working memory. *Proc Natl Acad Sci U S A* 112(2):E214–219. [PubMed: 25540412]

341. Brody CD, Hernandez A, Zainos A, & Romo R (2003) Timing and Neural Encoding of Somatosensory Parametric Working Memory in Macaque Prefrontal Cortex. *Cereb. Cortex* 13(11):1196–1207. [PubMed: 14576211]
342. Hernandez A, Zainos A, & Romo R (2002) Temporal evolution of a decision-making process in medial premotor cortex. *Neuron* 33(6):959–972. [PubMed: 11906701]
343. Vergara J, Rivera N, Rossi-Pool R, & Romo R (2016) A neural parametric code for storing information of more than one sensory modality in working memory. *Neuron*.
344. Constantinidis C (2016) A Code for Cross-Modal Working Memory. *Neuron* 89(1):3–5. [PubMed: 26748084]
345. Koch C & Fuster JM (1989) Unit activity in monkey parietal cortex related to haptic perception and temporary memory. *Experimental Brain Research* 76(2):292–306. [PubMed: 2767186]
346. Romo R, Hernandez A, Zainos A, Lemus L, & Brody CD (2002) Neuronal correlates of decision-making in secondary somatosensory cortex. *Nat Neurosci* 5(11):1217–1225. [PubMed: 12368806]
347. Romo R & Salinas E (2003) Flutter discrimination: neural codes, perception, memory and decision making. *Nat Rev Neurosci* 4(3):203–218. [PubMed: 12612633]
348. Zhou YD & Fuster JM (1996) Mnemonic neuronal activity in somatosensory cortex. *Proceedings of the National Academy of Sciences of the United States of America* 93(19):10533–10537. [PubMed: 8927629]
349. Jones EG, Dell'Anna ME, Molinari M, Rausell E, & Hashikawa T (1995) Subdivisions of macaque monkey auditory cortex revealed by calcium-binding protein immunoreactivity. *J Comp Neurol* 362(2):153–170. [PubMed: 8576431]
350. Kosaki H, Hashikawa T, He J, & Jones EG (1997) Tonotopic organization of auditory cortical fields delineated by parvalbumin immunoreactivity in macaque monkeys. *J Comp Neurol* 386(2):304–316. [PubMed: 9295154]
351. Kaas JH & Hackett TA (2000) Subdivisions of auditory cortex and processing streams in primates. *Proc Natl Acad Sci U S A* 97(22):11793–11799. [PubMed: 11050211]
352. Rauschecker JP, Tian B, & Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268(5207):111–114. [PubMed: 7701330]
353. Romanski LM, Bates JF, & Goldman-Rakic PS (1999) Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *Journal of Comparative Neurology* 403(2):141–157.
354. Romanski LM, et al. (1999) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience* 2(12):1131–1136. [PubMed: 10570492]
355. Hackett TA, Stepniewska I, & Kaas JH (1999) Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Research* 817(1-2):45–58. [PubMed: 9889315]
356. Plakke B, Ng CW, & Poremba A (2013) Neural correlates of auditory recognition memory in primate lateral prefrontal cortex. *Neuroscience* 244:62–76. [PubMed: 23583656]
357. Romanski LM & Goldman-Rakic PS (2002) An auditory domain in primate prefrontal cortex. *Nat Neurosci* 5(1):15–16. [PubMed: 11753413]
358. Bodner M, Kroger J, & Fuster JM (1996) Auditory memory cells in dorsolateral prefrontal cortex. *Neuroreport* 7(12):1905–1908. [PubMed: 8905689]
359. Fuster JM, Bodner M, & Kroger JK (2000) Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature* 405(6784):347–351. [PubMed: 10830963]
360. Rama P, et al. (2004) Dissociable functional cortical topographies for working memory maintenance of voice identity and location. *Cereb Cortex* 14(7):768–780. [PubMed: 15084491]
361. Sugihara T, Diltz MD, Averbeck BB, & Romanski LM (2006) Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. *J Neurosci* 26(43):11138–11147. [PubMed: 17065454]
362. Cohen YE, Hauser MD, & Russ BE (2006) Spontaneous processing of abstract categorical information in the ventrolateral prefrontal cortex. *Biol Lett* 2(2):261–265. [PubMed: 17148378]
363. Hwang J & Romanski LM (2015) Prefrontal neuronal responses during audiovisual mnemonic processing. *J Neurosci* 35(3):960–971. [PubMed: 25609614]

364. Scott BH, Mishkin M, & Yin P (2014) Neural correlates of auditory short-term memory in rostral superior temporal cortex. *Curr Biol* 24(23):2767–2775. [PubMed: 25456448]
365. Gottlieb Y, Vaadia E, & Abeles M (1989) Single unit activity in the auditory cortex of a monkey performing a short term memory task. *Exp Brain Res* 74(1):139–148. [PubMed: 2924831]
366. Bigelow J, Rossi B, & Poremba A (2014) Neural correlates of short-term memory in primate auditory cortex. *Front Neurosci* 8:250. [PubMed: 25177266]
367. Scott BH & Mishkin M (2016) Auditory short-term memory in the primate auditory cortex. *Brain Res* 1640(Pt B):264–277. [PubMed: 26541581]
368. Lemus L, Hernandez A, & Romo R (2009) Neural codes for perceptual discrimination of acoustic flutter in the primate auditory cortex. *Proc Natl Acad Sci U S A* 106(23):9471–9476. [PubMed: 19458263]
369. Plakke B, Hwang J, & Romanski LM (2015) Inactivation of Primate Prefrontal Cortex Impairs Auditory and Audiovisual Working Memory. *J Neurosci* 35(26):9666–9675. [PubMed: 26134649]
370. Cavada C, Company T, Tejedor J, Cruz-Rizzolo RJ, & Reinoso-Suarez F (2000) The anatomical connections of the macaque monkey orbitofrontal cortex. A review. *Cereb Cortex* 10(3):220–242. [PubMed: 10731218]
371. Carmichael ST, Clugnet MC, & Price JL (1994) Central olfactory connections in the macaque monkey. *J Comp Neurol* 346(3):403–434. [PubMed: 7527806]
372. Ifuku H, Hirata S, Nakamura T, & Ogawa H (2003) Neuronal activities in the monkey primary and higher-order gustatory cortices during a taste discrimination delayed GO/NOGO task and after reversal. *Neurosci Res* 47(2):161–175. [PubMed: 14512141]
373. Lara AH, Kennerley SW, & Wallis JD (2009) Encoding of gustatory working memory by orbitofrontal neurons. *J Neurosci* 29(3):765–774. [PubMed: 19158302]
374. Colflesh GJ & Conway AR (2007) Individual differences in working memory capacity and divided attention in dichotic listening. *Psychon Bull Rev* 14(4):699–703. [PubMed: 17972736]
375. Luck SJ & Vogel EK (1997) The capacity of visual working memory for features and conjunctions. *Nature* 390(6657):279–281. [PubMed: 9384378]
376. Unsworth N & Robison MK (2020) Working memory capacity and sustained attention: A cognitive-energetic perspective. *J Exp Psychol Learn Mem Cogn* 46(1):77–103. [PubMed: 30998072]
377. Simons DJ & Chabris CF (1999) Gorillas in our midst: sustained inattention blindness for dynamic events. *Perception* 28(9):1059–1074. [PubMed: 10694957]
378. Freeman J & Simoncelli EP (2011) Metamers of the ventral stream. *Nat Neurosci* 14(9):1195–1201. [PubMed: 21841776]
379. Cohen MA, Dennett DC, & Kanwisher N (2016) What is the Bandwidth of Perceptual Experience? *Trends Cogn Sci* 20(5):324–335. [PubMed: 27105668]
380. Pashler H (1988) Familiarity and visual change detection. *PerceptPsychophys* 44(4):369–378.
381. Rouders JN, Morey RD, Morey CC, & Cowan N (2011) How to measure working memory capacity in the change detection paradigm. *Psychon Bull Rev* 18(2):324–330. [PubMed: 21331668]
382. Foster JJ, Bsales EM, Jaffe RJ, & Awh E (2017) Alpha-Band Activity Reveals Spontaneous Representations of Spatial Position in Visual Working Memory. *Curr Biol* 27(20):3216–3223 e3216. [PubMed: 29033335]
383. Schurgin MW, Wixted JT, & Brady TF (2020) Psychophysical scaling reveals a unified theory of visual memory strength. *Nat Hum Behav* 4(11):1156–1172. [PubMed: 32895546]
384. Luria R, Sessa P, Gotler A, Jolicoeur P, & Dell'Acqua R (2010) Visual short-term memory capacity for simple and complex objects. *J Cogn Neurosci* 22(3):496–512. [PubMed: 19301998]
385. Luck SJ & Vogel EK (2013) Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends Cogn Sci* 17(8):391–400. [PubMed: 23850263]
386. Bays PM & Husain M (2008) Dynamic shifts of limited working memory resources in human vision. *Science* 321(5890):851–854. [PubMed: 18687968]

387. Doherty JM & Logie RH (2016) Resource-sharing in multiple-component working memory. *Mem Cognit* 44(8):1157–1167.
388. Bae GY & Luck SJ (2018) Dissociable Decoding of Spatial Attention and Working Memory from EEG Oscillations and Sustained Potentials. *J Neurosci* 38(2):409–422. [PubMed: 29167407]
389. Bae GY & Luck SJ (2019) Decoding motion direction using the topography of sustained ERPs and alpha oscillations. *Neuroimage* 184:242–255. [PubMed: 30223063]
390. Vogel EK & Machizawa MG (2004) Neural activity predicts individual differences in visual working memory capacity. *Nature* 428(6984):748–751. [PubMed: 15085132]
391. Schneegans S & Bays PM (2018) Drift in Neural Population Activity Causes Working Memory to Deteriorate Over Time. *J Neurosci* 38(21):4859–4869. [PubMed: 29703786]
392. Lisman JE & Idiart MA (1995) Storage of 7 +/- 2 short-term memories in oscillatory subcycles. *Science* 267(5203):1512–1515. [PubMed: 7878473]
393. Raffone A & Wolters G (2001) A cortical mechanism for binding in visual working memory. *J Cogn Neurosci* 13(6):766–785. [PubMed: 11564321]
394. Heyselaar E, Johnston K, & Pare M (2011) A change detection approach to study visual working memory of the macaque monkey. *J Vis* 11(3).
395. Lara AH & Wallis JD (2012) Capacity and precision in an animal model of visual short-term memory. *J Vis* 12(3).
396. Buschman TJ, Siegel M, Roy JE, & Miller EK (2011) Neural substrates of cognitive capacity limitations. *Proc Natl Acad Sci U S A* 108(27):11252–11255. [PubMed: 21690375]
397. Panichello MF, DePasquale B, Pillow JW, & Buschman TJ (2019) Error-correcting dynamics in visual working memory. *Nat Commun* 10(1):3366. [PubMed: 31358740]
398. Edin F, et al. (2009) Mechanism for top-down control of working memory capacity. *Proc Natl Acad Sci U S A* 106(16):6802–6807. [PubMed: 19339493]
399. Tang H, Qi XL, Riley MR, & Constantinidis C (2019) Working memory capacity is enhanced by distributed prefrontal activation and invariant temporal dynamics. *Proc Natl Acad Sci U S A* 116(14):7095–7100. [PubMed: 30877250]
400. Goelet P, Castellucci VF, Schacher S, & Kandel ER (1986) The long and the short of long-term memory--a molecular framework. *Nature* 322(6078):419–422. [PubMed: 2874497]
401. Brooks DN (1975) Long and short term memory in head injured patients. *Cortex* 11(4):329–340. [PubMed: 1222577]
402. Pertzov Y, Manohar S, & Husain M (2017) Rapid forgetting results from competition over time between items in visual working memory. *J Exp Psychol Learn Mem Cogn* 43(4):528–536. [PubMed: 27668485]
403. Sweller J (2016) Working Memory, Long-term Memory, and Instructional Design. *Journal of Applied Research in Memory and Cognition* 5(4):360–367.
404. Ruchkin DS, Grafman J, Cameron K, & Berndt RS (2003) Working memory retention systems: a state of activated long-term memory. *Behav Brain Sci* 26(6):709–728; discussion 728-777. [PubMed: 15377128]
405. Ranganath C, Johnson MK, & D'Esposito M (2003) Prefrontal activity associated with working memory and episodic long-term memory. *Neuropsychologia* 41(3):378–389. [PubMed: 12457762]
406. Cowan N (2005) Working Memory Capacity. *Psychol. Press* 260.
407. Kumar S, et al. (2016) A Brain System for Auditory Working Memory. *J Neurosci* 36(16):4492–4505. [PubMed: 27098693]
408. Chen T & Naveh-Benjamin M (2012) Assessing the associative deficit of older adults in long-term and short-term/working memory. *Psychol Aging* 27(3):666–682. [PubMed: 22308997]
409. Hulme C, Maughan S, Brown G (1991) Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language* 30(6):685–701.
410. Dash PK, Moore AN, Kobori N, & Runyan JD (2007) Molecular activity underlying working memory. *Learn Mem* 14(8):554–563. [PubMed: 17690339]

411. Qi XL, Meyer T, Stanford TR, & Constantinidis C (2012) Neural correlates of a decision variable before learning to perform a match/non-match task. *J Neurosci* 32(18):6161–6169. [PubMed: 22553022]
412. Lisman JE & Grace AA (2005) The hippocampal-VTA loop: controlling the entry of information into long-term memory. *Neuron* 46(5):703–713. [PubMed: 15924857]
413. Zhao D, Zhou YD, Bodner M, & Ku Y (2018) The Causal Role of the Prefrontal Cortex and Somatosensory Cortex in Tactile Working Memory. *Cereb Cortex* 28(10):3468–3477. [PubMed: 28968894]
414. Hannula H, et al. (2010) Increasing top-down suppression from prefrontal cortex facilitates tactile working memory. *Neuroimage* 49(1):1091–1098. [PubMed: 19643184]
415. Gallace A & Spence C (2009) The cognitive and neural correlates of tactile memory. *Psychol Bull* 135(3):380–406. [PubMed: 19379022]
416. Calvert GA, et al. (1997) Activation of auditory cortex during silent lipreading. *Science* 276(5312):593–596. [PubMed: 9110978]
417. Blake R, Cepeda NJ, & Hiris E (1997) Memory for visual motion. *J Exp Psychol Hum Percept Perform* 23(2):353–369. [PubMed: 9103999]
418. Magnussen S & Greenlee MW (1992) Retention and disruption of motion information in visual short-term memory. *J Exp Psychol Learn Mem Cogn* 18(1):151–156. [PubMed: 1532017]
419. Magnussen S, Idas E, & Myhre SH (1998) Representation of orientation and spatial frequency in perception and memory: a choice reaction-time analysis. *J Exp Psychol Hum Percept Perform* 24(3):707–718. [PubMed: 9627410]
420. Cowan N & AuBuchon AM (2008) Short-term memory loss over time without retroactive stimulus interference. *Psychon Bull Rev* 15(1):230–235. [PubMed: 18605508]
421. Oberauer K (2019) Is Rehearsal an Effective Maintenance Strategy for Working Memory? *Trends Cogn Sci* 23(9):798–809. [PubMed: 31301953]
422. Pinal D, Zurrón M, & Diaz F (2014) Effects of load and maintenance duration on the time course of information encoding and retrieval in working memory: from perceptual analysis to post-categorization processes. *Front Hum Neurosci* 8:165. [PubMed: 24744715]
423. Wixted JT (1991) Conditions and consequences of maintenance rehearsal. *J Exp Psychol Learn Mem Cogn* 17(5):963–973. [PubMed: 1834777]
424. Sperling G (1960) The information available in brief visual presentations (American Psychological Association, Washington) p 29 p.
425. Portrat S, Barrouillet P, & Camos V (2008) Time-related decay or interference-based forgetting in working memory? *J Exp Psychol Learn Mem Cogn* 34(6):1561–1564. [PubMed: 18980415]
426. Elliott LA & Strawhorn RJ (1976) Interference in short-term memory from vocalization: aural versus visual modality differences. *J Exp Psychol Hum Learn* 2(6):705–711. [PubMed: 1010993]
427. Wickelgren WA (1965) Acoustic Similarity and Intrusion Errors in Short-Term Memory. *J Exp Psychol* 70:102–108. [PubMed: 14315119]
428. Proctor R, Fagnani C (1978) Effects of distractor-stimulus modality in the Brown–Peterson distractor task. *Journal of Experimental Psychology: Human Learning and Memory* 4(6):676–684.
429. Balan PF & Ferrera VP (2003) Effects of spontaneous eye movements on spatial memory in macaque periarculate cortex. *J Neurosci* 23(36):11392–11401. [PubMed: 14673003]
430. Papadimitriou C, Holmes CD, & Snyder LH (2021) Primate Spatial Memory Cells Become Tuned Early and Lose Tuning at Cell-Specific Times. *Cereb Cortex*.
431. Holmes CD, Papadimitriou C, & Snyder LH (2018) Dissociation of LFP Power and Tuning in the Frontal Cortex during Memory. *J Neurosci* 38(38):8177–8186. [PubMed: 30093534]
432. White JM, Sparks DL, & Stanford TR (1994) Saccades to remembered target locations: an analysis of systematic and variable errors. *Vision Res* 34(1):79–92. [PubMed: 8116271]
433. Lowet E, et al. (2018) Enhanced Neural Processing by Covert Attention only during Microsaccades Directed toward the Attended Stimulus. *Neuron* 99(1):207–214 e203. [PubMed: 29937279]

434. D'Esposito M, Ballard D, Zarahn E, & Aguirre GK (2000) The role of prefrontal cortex in sensory memory and motor preparation: an event-related fMRI study. *Neuroimage* 11(5 Pt 1):400–408. [PubMed: 10806027]
435. Laubach M, Amarante LM, Swanson K, & White SR (2018) What, If Anything, Is Rodent Prefrontal Cortex? *eNeuro* 5(5).
436. Erlich JC, Bialek M, & Brody CD (2011) A cortical substrate for memory-guided orienting in the rat. *Neuron* 72(2):330–343. [PubMed: 22017991]
437. Esmaili V & Diamond ME (2019) Neuronal Correlates of Tactile Working Memory in Prefrontal and Vibrissal Somatosensory Cortex. *Cell Rep* 27(11):3167–3181 e3165. [PubMed: 31189103]
438. Kamigaki T & Dan Y (2017) Delay activity of specific prefrontal interneuron subtypes modulates memory-guided behavior. *Nat Neurosci* 20(6):854–863. [PubMed: 28436982]
439. Scott BB, et al. (2017) Fronto-parietal Cortical Circuits Encode Accumulated Evidence with a Diversity of Timescales. *Neuron* 95(2):385–398 e385. [PubMed: 28669543]
440. Baeg EH, et al. (2003) Dynamics of population code for working memory in the prefrontal cortex. *Neuron* 40(1):177–188. [PubMed: 14527442]
441. Bolkan SS, et al. (2017) Thalamic projections sustain prefrontal activity during working memory maintenance. *Nat Neurosci* 20(7):987–996. [PubMed: 28481349]
442. Guo K, Yamawaki N, Svoboda K, & Shepherd GMG (2018) Anterolateral Motor Cortex Connects with a Medial Subdivision of Ventromedial Thalamus through Cell Type-Specific Circuits, Forming an Excitatory Thalamo-Cortico-Thalamic Loop via Layer 1 Apical Tuft Dendrites of Layer 5B Pyramidal Tract Type Neurons. *J Neurosci* 38(41):8787–8797. [PubMed: 30143573]
443. Lyamzin D & Benucci A (2019) The mouse posterior parietal cortex: Anatomy and functions. *Neurosci Res* 140:14–22. [PubMed: 30465783]
444. Kahn JB, et al. (2012) Medial prefrontal lesions in mice impair sustained attention but spare maintenance of information in working memory. *Learn Mem* 19(11):513–517. [PubMed: 23073640]
445. Muir JL, Everitt BJ, & Robbins TW (1996) The cerebral cortex of the rat and visual attentional function: dissociable effects of mediofrontal, cingulate, anterior dorsolateral, and parietal cortex lesions on a five-choice serial reaction time task. *Cereb Cortex* 6(3):470–481. [PubMed: 8670672]
446. Granon S, Hardouin J, Courtier A, & Poucet B (1998) Evidence for the involvement of the rat prefrontal cortex in sustained attention. *Q J Exp Psychol B* 51(3):219–233. [PubMed: 9786783]
447. Preuss TM (1995) Do rats have prefrontal cortex? The rose-woolsey-akert program reconsidered. *J Cogn Neurosci* 7(1):1–24. [PubMed: 23961750]
448. Balakhonov D & Rose J (2017) Crows Rival Monkeys in Cognitive Capacity. *Sci Rep* 7(1):8809. [PubMed: 28821812]
449. Veit L & Nieder A (2013) Abstract rule neurons in the endbrain support intelligent behaviour in corvid songbirds. *Nat Commun* 4:2878. [PubMed: 24285080]
450. Gunturkun O (2005) The avian 'prefrontal cortex' and cognition. *Curr Opin Neurobiol* 15(6):686–693. [PubMed: 16263260]
451. Emery NJ & Clayton NS (2005) Evolution of the avian brain and intelligence. *Curr Biol* 15(23):R946–950. [PubMed: 16332520]
452. Kirsch JA, Gunturkun O, & Rose J (2008) Insight without cortex: lessons from the avian brain. *Conscious Cogn* 17(2):475–483. [PubMed: 18440242]
453. Reiner A, et al. (2004) Revised nomenclature for avian telencephalon and some related brainstem nuclei. *J Comp Neurol* 473(3):377–414. [PubMed: 15116397]
454. Shanahan M, Bingman VP, Shimizu T, Wild M, & Gunturkun O (2013) Large-scale network organization in the avian forebrain: a connectivity matrix and theoretical analysis. *Front Comput Neurosci* 7:89. [PubMed: 23847525]
455. Emery NJ & Clayton NS (2004) The mentality of crows: convergent evolution of intelligence in corvids and apes. *Science* 306(5703):1903–1907. [PubMed: 15591194]
456. Butler AB & Cotterill RM (2006) Mammalian and avian neuroanatomy and the question of consciousness in birds. *Biol Bull* 211(2):106–127. [PubMed: 17062871]

457. Rinnert P, Kirschhock ME, & Nieder A (2019) Neuronal Correlates of Spatial Working Memory in the Endbrain of Crows. *Curr Biol* 29(16):2616–2624 e2614. [PubMed: 31378607]
458. Diekamp B, Kalt T, & Gunturkun O (2002) Working memory neurons in pigeons. *J Neurosci* 22(4):RC210. [PubMed: 11844844]
459. Kalt T, Diekamp B, & Gunturkun O (1999) Single unit activity during a Go/NoGo task in the "prefrontal cortex" of pigeons. *Brain Res* 839(2):263–278. [PubMed: 10519049]
460. Veit L, Hartmann K, & Nieder A (2014) Neuronal correlates of visual working memory in the corvid endbrain. *J Neurosci* 34(23):7778–7786. [PubMed: 24899702]
461. Kroner S & Gunturkun O (1999) Afferent and efferent connections of the caudolateral neostriatum in the pigeon (*Columba livia*): a retro- and anterograde pathway tracing study. *J Comp Neurol* 407(2):228–260. [PubMed: 10213093]
462. Diekamp B, Gagliardo A, & Gunturkun O (2002) Nonspatial and subdivision-specific working memory deficits after selective lesions of the avian prefrontal cortex. *J Neurosci* 22(21):9573–9580. [PubMed: 12417682]
463. Johnston M, Porter B, & Colombo M (2019) Delay activity in pigeon nidopallium caudolaterale during a variable-delay memory task. *Behav Neurosci* 133(6):563–568. [PubMed: 31464458]
464. Johnston M, Anderson C, & Colombo M (2017) Neural correlates of sample-coding and reward-coding in the delay activity of neurons in the entopallium and nidopallium caudolaterale of pigeons (*Columba livia*). *Behav Brain Res* 317:382–392. [PubMed: 27720742]
465. Yakovlev PI & Lecours AR eds (1967) *The myelogenetic cycles of regional maturation of the brain* (Blackwell, Oxford), pp 3–70.
466. Pfefferbaum A, et al. (1994) A quantitative magnetic resonance imaging study of changes in brain morphology from infancy to late adulthood. *Arch Neurol* 51(9):874–887. [PubMed: 8080387]
467. Jernigan TL, et al. (1991) Cerebral structure on MRI, Part I: Localization of age-related changes. *Biol Psychiatry* 29(1):55–67. [PubMed: 2001446]
468. Sowell ER, Delis D, Stiles J, & Jernigan TL (2001) Improved memory functioning and frontal lobe maturation between childhood and adolescence: a structural MRI study. *J Int Neuropsychol Soc* 7(3):312–322. [PubMed: 11311032]
469. Chugani HT, Phelps ME, & Mazziotta JC (1987) Positron emission tomography study of human brain functional development. *Ann Neurol* 22(4):487–497. [PubMed: 3501693]
470. Gathercole SE, Pickering SJ, Ambridge B, & Wearing H (2004) The structure of working memory from 4 to 15 years of age. *Dev Psychol* 40(2):177–190. [PubMed: 14979759]
471. Fry AF & Hale S (2000) Relationships among processing speed, working memory, and fluid intelligence in children. *Biol Psychol* 54(1-3):1–34. [PubMed: 11035218]
472. Spronk M, Vogel EK, & Jonkman LM (2012) Electrophysiological evidence for immature processing capacity and filtering in visuospatial working memory in adolescents. *PLoS One* 7(8):e42262. [PubMed: 22927923]
473. Luna B, Garver KE, Urban TA, Lazar NA, & Sweeney JA (2004) Maturation of cognitive processes from late childhood to adulthood. *Child Dev* 75(5):1357–1372. [PubMed: 15369519]
474. Crone EA, Wendelken C, Donohue S, van Leijenhorst L, & Bunge SA (2006) Neurocognitive development of the ability to manipulate information in working memory. *Proc Natl Acad Sci U S A* 103(24):9315–9320. [PubMed: 16738055]
475. Isbell E, Fukuda K, Neville HJ, & Vogel EK (2015) Visual working memory continues to develop through adolescence. *Front Psychol* 6:696. [PubMed: 26074849]
476. Alloway TP, Gathercole SE, & Pickering SJ (2006) Verbal and visuospatial short-term and working memory in children: are they separable? *Child Dev* 77(6):1698–1716. [PubMed: 17107455]
477. Montez D, Calabro F, & Luna B (2017) The expression of established cognitive brain states stabilizes with working memory development. *eLife Epub ahead of print*.
478. Lee EY, et al. (2010) Visual working memory deficits in patients with Parkinson's disease are due to both reduced storage capacity and impaired ability to filter out irrelevant information. *Brain* 133(9):2677–2689. [PubMed: 20688815]

479. Westerberg H, Hirvikoski T, Forssberg H, & Klingberg T (2004) Visuo-spatial working memory span: a sensitive measure of cognitive deficits in children with ADHD. *Child Neuropsychol* 10(3):155–161. [PubMed: 15590494]
480. Park S, Holzman PS, & Goldman-Rakic PS (1995) Spatial working memory deficits in the relatives of schizophrenic patients. *Archives of General Psychiatry* 52(10):821–828. [PubMed: 7575101]
481. Pantelis C, et al. (1997) Frontal-striatal cognitive deficits in patients with chronic schizophrenia. *Brain* 120(Pt 10):1823–1843. [PubMed: 9365373]
482. Burgess GC, et al. (2010) Attentional control activation relates to working memory in attention-deficit/hyperactivity disorder. *Biol Psychiatry* 67(7):632–640. [PubMed: 20060961]
483. Bunge SA, Dudukovic NM, Thomason ME, Vaidya CJ, & Gabrieli JD (2002) Immature frontal lobe contributions to cognitive control in children: evidence from fMRI. *Neuron* 33(2):301–311. [PubMed: 11804576]
484. Olesen PJ, Nagy Z, Westerberg H, & Klingberg T (2003) Combined analysis of DTI and fMRI data reveals a joint maturation of white and grey matter in a fronto-parietal network. *Brain Res Cogn Brain Res* 18(1):48–57. [PubMed: 14659496]
485. Olesen PJ, Macoveanu J, Tegner J, & Klingberg T (2007) Brain activity related to working memory and distraction in children and adults. *Cereb Cortex* 17(5):1047–1054. [PubMed: 16801377]
486. Burgund ED, Lugar HM, Miezin FM, Schlaggar BL, & Petersen SE (2006) The development of sustained and transient neural activity. *Neuroimage* 29(3):812–821. [PubMed: 16236529]
487. Klingberg T, Forssberg H, & Westerberg H (2002) Training of working memory in children with ADHD. *J Clin Exp Neuropsychol* 24(6):781–791. [PubMed: 12424652]
488. Kwon H, Reiss AL, & Menon V (2002) Neural basis of protracted developmental changes in visuo-spatial working memory. *Proc Natl Acad Sci U S A* 99(20):13336–13341. [PubMed: 12244209]
489. Simmonds DJ, Hallquist MN, & Luna B (2017) Protracted development of executive and mnemonic brain systems underlying working memory in adolescence: a longitudinal fMRI study. *Neuroimage*.
490. Plant TM, Ramaswamy S, Simorangkir D, & Marshall GR (2005) Postnatal and pubertal development of the rhesus monkey (*Macaca mulatta*) testis. *Ann N Y Acad Sci* 1061:149–162. [PubMed: 16467264]
491. Herman RA, Zehr JL, & Wallen K (2006) Prenatal androgen blockade accelerates pubertal development in male rhesus monkeys. *Psychoneuroendocrinology* 31(1):118–130. [PubMed: 16112492]
492. Fuster JM (2002) Frontal lobe and cognitive development. *J Neurocytol* 31(3-5):373–385. [PubMed: 12815254]
493. Lewis DA (1997) Development of the prefrontal cortex during adolescence: insights into vulnerable neural circuits in schizophrenia. *Neuropsychopharmacology* 16(6):385–398. [PubMed: 9165494]
494. Malkova L, Heuer E, & Saunders RC (2006) Longitudinal magnetic resonance imaging study of rhesus monkey brain development. *Eur J Neurosci* 24(11):3204–3212. [PubMed: 17156381]
495. Hoftman GD & Lewis DA (2011) Postnatal developmental trajectories of neural circuits in the primate prefrontal cortex: identifying sensitive periods for vulnerability to schizophrenia. *Schizophr Bull* 37(3):493–503. [PubMed: 21505116]
496. Zhou X, et al. (2016) Neural correlates of working memory development in adolescent primates. *Nat Commun* 7:13423. [PubMed: 27827365]
497. Montez DF, Calabro FJ, & Luna B (2019) Working memory improves developmentally as neural processes stabilize. *PLoS One* 14(3):e0213010. [PubMed: 30845260]
498. Zhou X, et al. (2014) Age-dependent changes in prefrontal intrinsic connectivity. *Proc Natl Acad Sci U S A* 111(10):3853–3858. [PubMed: 24567390]
499. Gonzalez-Burgos G, et al. (2015) Functional Maturation of GABA Synapses During Postnatal Development of the Monkey Dorsolateral Prefrontal Cortex. *Cereb Cortex* 25(11):4076–4093. [PubMed: 24904071]

500. Klingberg T, Forssberg H, & Westerberg H (2002) Training of working memory in children with ADHD. *J. Clin. Exp. Neuropsychol* 24(6):781–791. [PubMed: 12424652]
501. Jaeggi SM, Buschkuhl M, Jonides J, & Perrig WJ (2008) Improving fluid intelligence with training on working memory. *Proc.Natl.Acad.Sci. U.S.A* 105(19):6829–6833. [PubMed: 18443283]
502. Westerberg H, et al. (2007) Computerized working memory training after stroke--a pilot study. *Brain Inj* 21(1)21–29. [PubMed: 17364516]
503. Subramaniam K, et al. (2012) Computerized cognitive training restores neural activity within the reality monitoring network in schizophrenia. *Neuron* 73(4):842–853. [PubMed: 22365555]
504. Anguera JA, et al. (2013) Video game training enhances cognitive control in older adults. *Nature* 501(7465):97–101. [PubMed: 24005416]
505. Cortese S, et al. (2015) Cognitive training for attention-deficit/hyperactivity disorder: meta-analysis of clinical and neuropsychological outcomes from randomized controlled trials. *Journal of the American Academy of Child and Adolescent Psychiatry* 54(3):164–174. [PubMed: 25721181]
506. Schwaighofer M, Fischer F, & Buhner M (2015) Does Working Memory Training Transfer? A Meta-Analysis Including Training Conditions as Moderators. *Educational Psychologist* 50(2):138–166.
507. Owen AM, et al. (2010) Putting brain training to the test. *Nature* 465(7299):775–778. [PubMed: 20407435]
508. Garavan H, Kelley D, Rosen A, Rao SM, & Stein EA (2000) Practice-related functional activation changes in a working memory task. *Microsc.Res.Tech* 51(1):54–63. [PubMed: 11002353]
509. Hempel A, et al. (2004) Plasticity of cortical activation related to working memory during training. *Am.J Psychiatry* 161(4):745–747. [PubMed: 15056524]
510. Dahlin E, Neely AS, Larsson A, Backman L, & Nyberg L (2008) Transfer of learning after updating training mediated by the striatum. *Science* 320(5882):1510–1512. [PubMed: 18556560]
511. Jolles DD, van Buchem MA, Crone EA, & Rombouts SA (2013) Functional brain connectivity at rest changes after working memory training. *Hum Brain Mapp* 34(2)396–406. [PubMed: 22076823]
512. Olesen PJ, Westerberg H, & Klingberg T (2004) Increased prefrontal and parietal activity after training of working memory. *Nat Neurosci* 7(1):75–79. [PubMed: 14699419]
513. Schneiders JA, Opitz B, Krick CM, & Mecklinger A (2011) Separating intra-modal and across-modal training effects in visual working memory: an fMRI investigation. *Cereb Cortex* 21(11):2555–2564. [PubMed: 21471559]
514. Kuhn S, et al. (2013) The dynamics of change in striatal activity following updating training. *Hum Brain Mapp* 34(7):1530–1541. [PubMed: 22331673]
515. Schweizer S, Grahn J, Hampshire A, Mobbs D, & Dalgleish T (2013) Training the emotional brain: improving affective control through emotional working memory training. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33(12):5301–5311. [PubMed: 23516294]
516. Takeuchi H, et al. (2013) Effects of working memory training on functional connectivity and cerebral blood flow during rest. *Cortex; a journal devoted to the study of the nervous system and behavior* 49(8):2106–2125. [PubMed: 23079491]
517. Mendoza-Halliday D & Martinez-Trujillo JC (2017) Neuronal population coding of perceived and memorized visual features in the lateral prefrontal cortex. *Nat Commun* 8:15471. [PubMed: 28569756]
518. Moreno-Bote R, et al. (2014) Information-limiting correlations. *Nat Neurosci* 17(10):1410–1417. [PubMed: 25195105]
519. Qi XL & Constantinidis C (2012) Correlated discharges in the primate prefrontal cortex before and after working memory training *European Journal of Neuroscience* 36(11)3538–3548.
520. Kobak D, et al. (2014) Demixed principal component analysis of population activity in higher cortical areas reveals independent representation of task parameters. *arXiv:1410.6031*.

521. Sarma A, Masse NY, Wang XJ, & Freedman DJ (2016) Task-specific versus generalized mnemonic representations in parietal and prefrontal cortices. *Nat Neurosci* 19(1):143–149. [PubMed: 26595652]
522. Asaad WF, Rainer G, & Miller EK (1998) Neural activity in the primate prefrontal cortex during associative learning. *Neuron* 21(6):1399–1407. [PubMed: 9883732]
523. Watanabe K & Funahashi S (2014) Neural mechanisms of dual-task interference and cognitive capacity limitation in the prefrontal cortex. *Nat Neurosci* 17(4):601–611. [PubMed: 24584049]
524. Kuboshima-Amemori S & Sawaguchi T (2007) Plasticity of the primate prefrontal cortex. *Neuroscientist* 13(3):229–240. [PubMed: 17519366]
525. Garcia-Cabezas MA, Joyce MKP, John YJ, Zikopoulos B, & Barbas H (2017) Mirror trends of plasticity and stability indicators in primate prefrontal cortex. *Eur J Neurosci* 46(8):2392–2405. [PubMed: 28921934]
526. McEwen BS & Morrison JH (2013) The brain on stress: vulnerability and plasticity of the prefrontal cortex over the life course. *Neuron* 79(1):16–29. [PubMed: 23849196]
527. Laroche S, Davis S, & Jay TM (2000) Plasticity at hippocampal to prefrontal cortex synapses: dual roles in working memory and consolidation. *Hippocampus* 10(4):438–446. [PubMed: 10985283]
528. Hempel CM, Hartman KH, Wang XJ, Turrigiano GG, & Nelson SB (2000) Multiple forms of short-term plasticity at excitatory synapses in rat medial prefrontal cortex. *J Neurophysiol* 83(5):3031–3041. [PubMed: 10805698]
529. Klingberg T, et al. (2005) Computerized training of working memory in children with ADHD--a randomized, controlled trial. *J Am Acad Child Adolesc Psychiatry* 44(2):177–186. [PubMed: 15689731]
530. Jaeggi SM, Buschkuhl M, Jonides J, & Perrig WJ (2008) Improving fluid intelligence with training on working memory. *Proc Natl Acad Sci U S A* 105(19):6829–6833. [PubMed: 18443283]
531. McNab F, et al. (2009) Changes in cortical dopamine D1 receptor binding associated with cognitive training. *Science* 323(5915):800–802. [PubMed: 19197069]
532. Sawaguchi T & Goldman-Rakic PS (1991) D1 dopamine receptors in prefrontal cortex: involvement in working memory. *Science* 251(4996):947–950. [PubMed: 1825731]
533. Backman L, et al. (2011) Effects of working-memory training on striatal dopamine release. *Science* 333(6043):718. [PubMed: 21817043]
534. Yokokura M, et al. (2020) In vivo imaging of dopamine D1 receptor and activated microglia in attention-deficit/hyperactivity disorder: a positron emission tomography study. *Mol Psychiatry*.
535. Arce E, et al. (2019) A novel approach to evaluate the pharmacodynamics of a selective dopamine D1/D5 receptor partial agonist (PF-06412562) in patients with stable schizophrenia. *J Psychopharmacol* 33(10):1237–1247. [PubMed: 31264510]
536. Gibson EM, et al. (2014) Neuronal activity promotes oligodendrogenesis and adaptive myelination in the mammalian brain. *Science* 344(6183):1252304. [PubMed: 24727982]
537. Yeung MS, et al. (2014) Dynamics of oligodendrocyte generation and myelination in the human brain. *Cell* 159(4):766–774. [PubMed: 25417154]
538. Takeuchi H, et al. (2010) Training of working memory impacts structural connectivity. *J Neurosci* 30(9):3297–3303. [PubMed: 20203189]
539. Bathelt J, Gathercole SE, Johnson A, & Astle DE (2018) Differences in brain morphology and working memory capacity across childhood. *Dev Sci* 21(3):e12579. [PubMed: 28748537]
540. Thompson TW, Waskom ML, & Gabrieli JD (2016) Intensive Working Memory Training Produces Functional Changes in Large-scale Frontoparietal Networks. *J Cogn Neurosci* 28(4):575–588. [PubMed: 26741799]
541. Chase W & Erickson KA (1982) Skill and Working Memory. *The Psychology of Learning and Motivation* 16.
542. Kliegl RJ, Heckhausen S, Baltes J, P. (1987) Mnemonic Training for the Acquisition of Skilled Digit Memory. *Cognition and Instruction* 4(4).

543. Verhaeghen P & Marcoen A (1996) On the mechanisms of plasticity in young and older adults after instruction in the method of loci: evidence for an amplification model. *Psychol Aging* 11(1):164–178. [PubMed: 8726382]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Didactic Synopsis

Working memory is a core component of higher cognitive functions. An empirical understanding of its neural basis is a key factor in unraveling the operations of the conscious mind and the origins of human thought.

Major Teaching Points:

- Persistent discharges in a network of prefrontal neurons connected with neurons in other brain areas represent information in working memory.
- This circuit behaves as a continuous attractor that can maintain a peak of activation, representing a stimulus.
- Synaptic mechanisms allow representation of latent information even without persistent activity; they influence behavior inasmuch as they influence the generation of spiking activity.
- The limits of WM capacity are attributed to the ability of the network to maintain simultaneous peaks of activity representing stimuli, without suppressing each other.
- The duration of WM is limited in the order of seconds. Drift and interference rather than decay explain most of the inaccuracy that accumulates in time.
- Persistent discharges appear in other vertebrate species including rodents and corvids.
- Improvements of working memory ability in childhood and adolescence parallel increases in persistent activity. Training in adulthood can similarly enhance WM through underlying changes in the persistent activity.

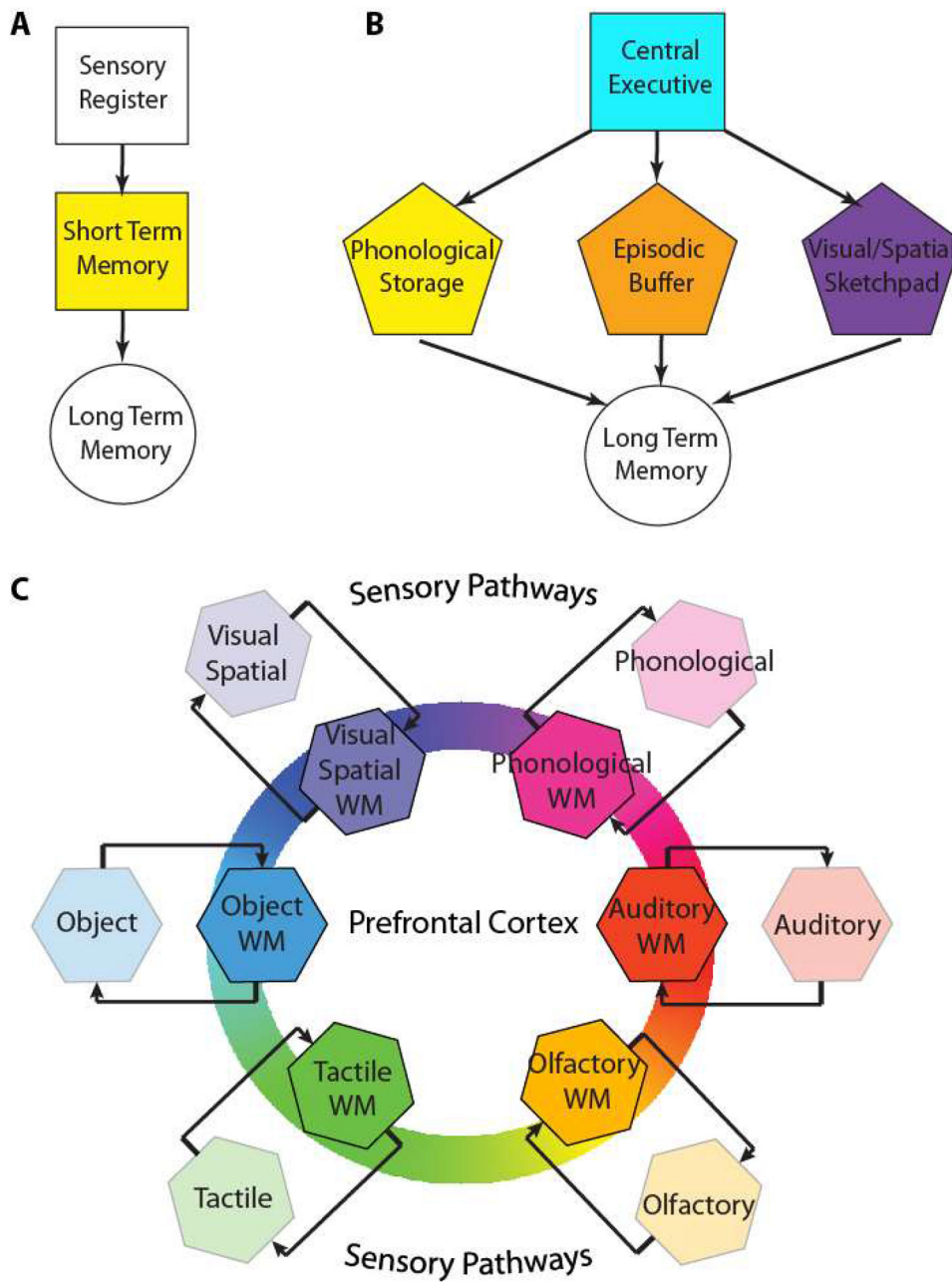


Figure 1. A. Schematic diagram of (A) the Atkinson and Shiffrin model, (B) Baddeley model, and (C) the model we are putting forth here (Jaffe and Constantinidis model). Colored squares in all schematic diagrams represent short/term working memory stages. The prefrontal cortex plays a central role in the maintenance of working memory through activity that reverberates in short- and long-distance loops including the later stages of the sensory pathways that originally transmit sensory information to working memory.

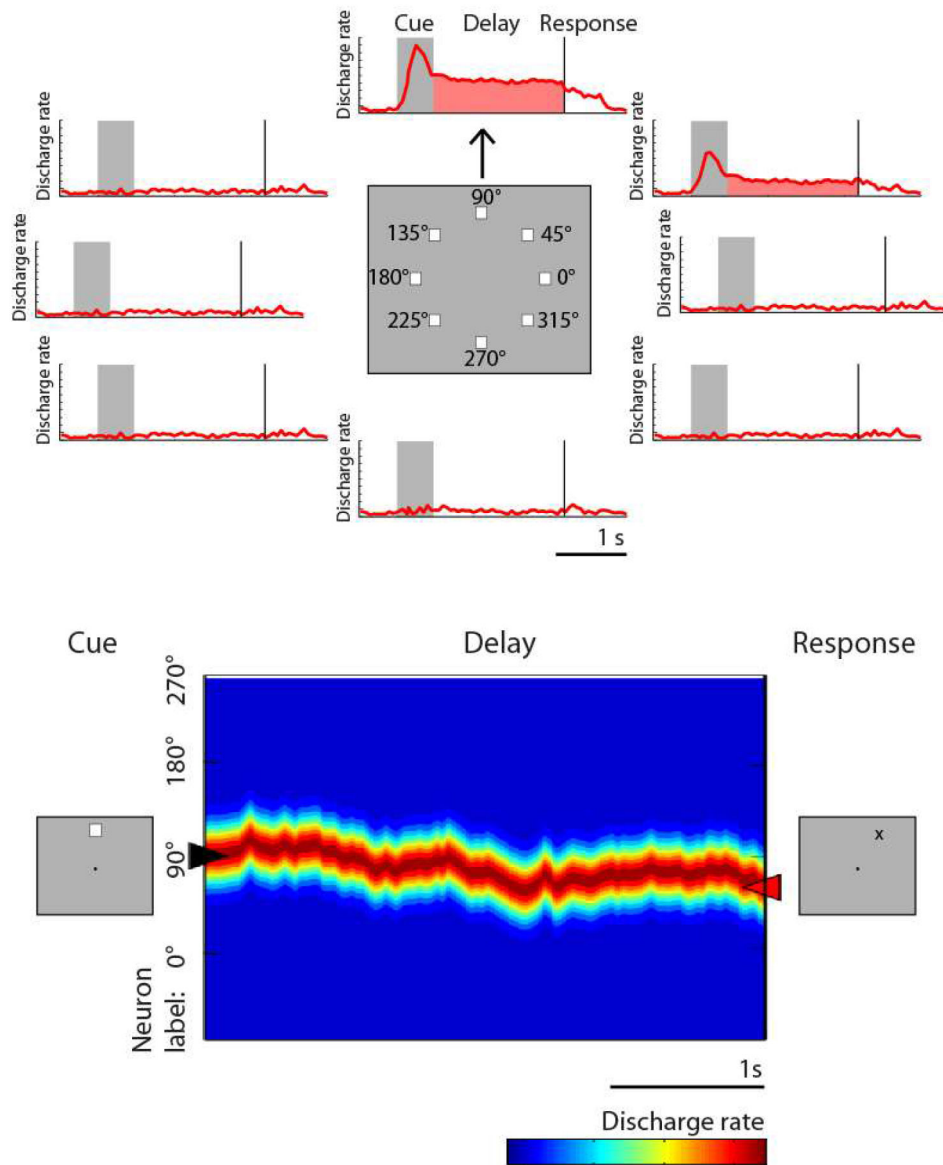


Figure 2. From single neuron responses to the bump attractor. Top, schematic illustration of responses of a single neuron to the ODR task for spatial working memory of a stimulus that appears at 8 different locations. Bottom, the population of neurons represent the stimulus location by the bump of activity in the network. Drifts of this activity result in errors. Adapted from Klingberg and Constantinidis, 2016.

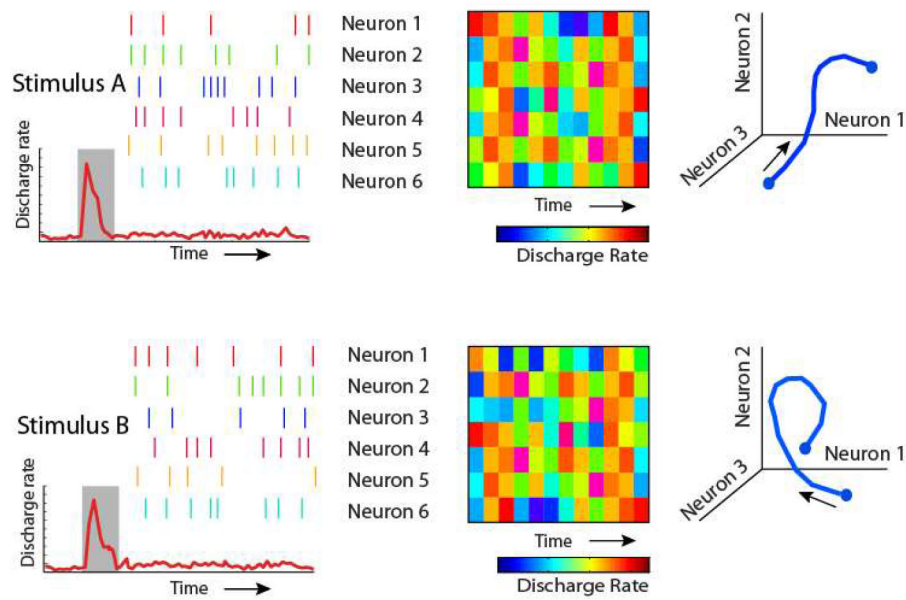


Figure 3. Schematic illustration of types of dynamic coding. Two different stimuli may not elicit an overall increase in firing rate during the delay period of working memory tasks (left panels). Their identity may still be decoded based on the pattern of spikes of individual neurons, which may be separable for different stimuli (middle panels). Furthermore, the pattern of activity of individual neurons may evolve dynamically in time, i.e. different neurons reach maximum activity at different time points during the trial, resulting in different trajectories in state space (right panels).

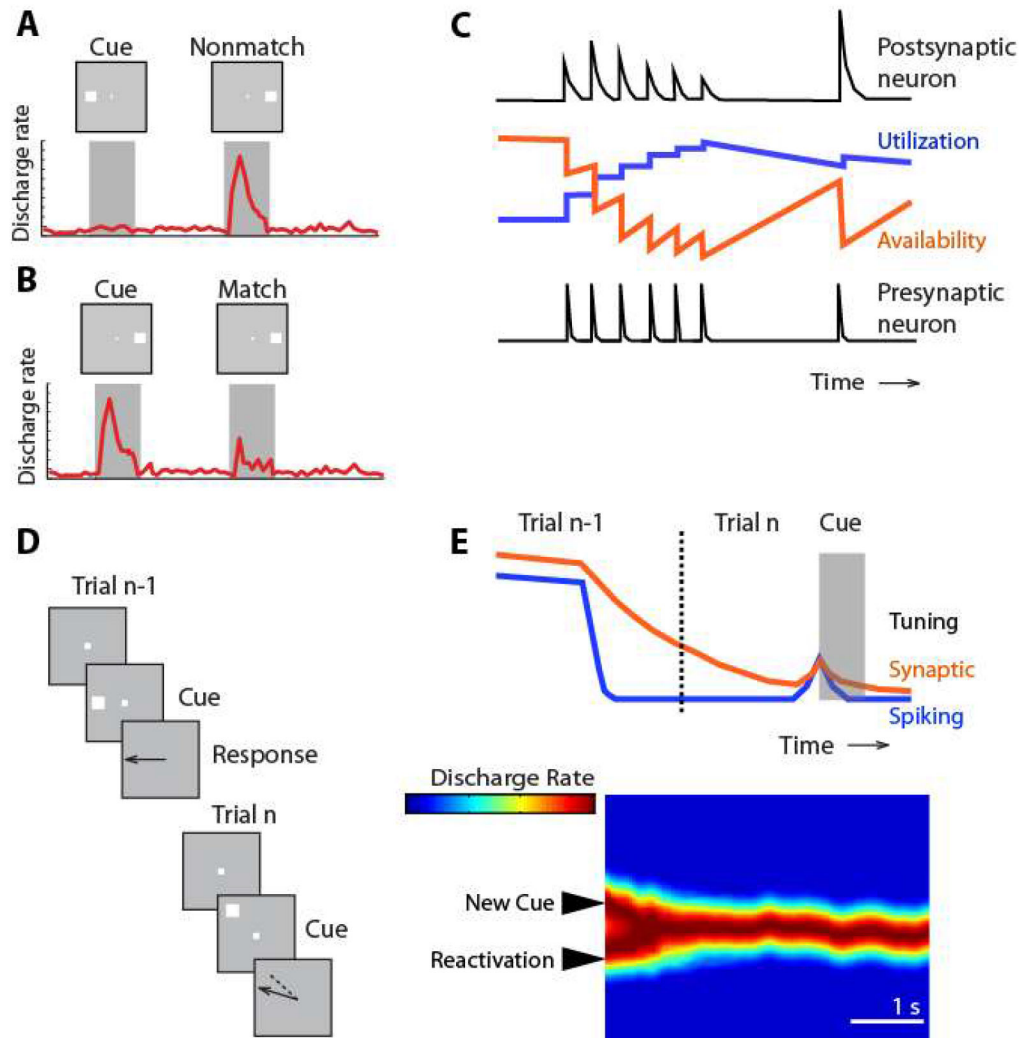


Figure 4. **A-B.** Schematic demonstration of the phenomenon of match suppression: a modulation of a stimulus firing rate depending on a previous stimulus, which does not depend on persistent spiking over the time interval between the two stimuli. **C.** The synaptic model of working memory. Utilization and availability of synaptic resources (e.g. Calcium concentration) modulate the magnitude of the postsynaptic potential generated by spikes elicited when a stimulus is presented for the first time, and when the same stimulus is presented for the second time. **D.** Schematic illustration of serial bias in working memory. The location of the previous stimulus influences the memory of the current stimulus location. **E.** During the inter-trial interval (dotted line), firing rate no longer represents the previous stimulus and the neuron exhibits no selectivity for the location of the preceding stimulus. However, this information continues to be maintained by synaptic mechanisms and when the next trial begins, selectivity for the preceding stimulus re-emerges. The interaction of this reactivated bump and the bump of activity caused by the cue appearance causes a slight deviation (bias) which is evident in the pattern of behavioral performance. Panel C, adapted from Mongillo et al., 2008; panel E from Barbosa et al., 2020.

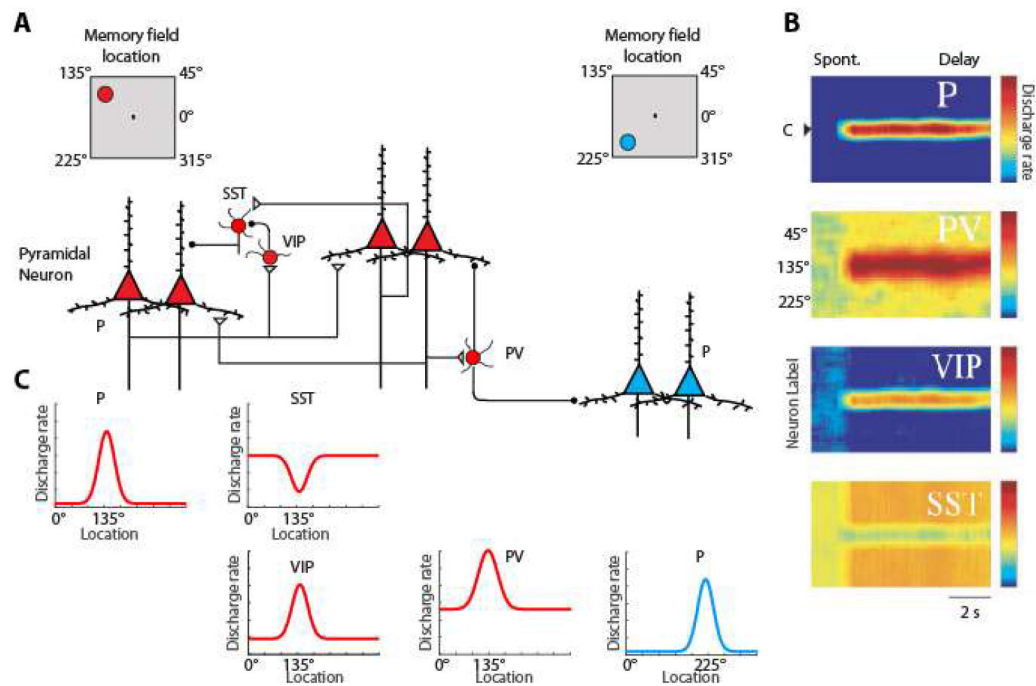


Figure 5. Division of labor model. Selectivity of Pyramidal Neurons (P) and three types of interneurons: Parvalbumin (PV), Calretinin (CR/VIP) and Calbindin (CB/SST) are shown. **A.** Insets on top are meant to illustrate that red-colored neurons on the left side of the figure are driven by a stimulus at the upper left of the screen, the 135° location, whereas blue-colored neurons on the right side of the figure are maximally activated by a stimulus in the lower left, 225° location. Excitatory synapses connect pyramidal neurons with similar preferences in the delay period that follows a stimulus in the upper left. **B.** Heat maps representing the activity of different neurons are plotted by preference for stimulus location (y-axis), as a function of time (x-axis). Adapted from Li et al., 2020.

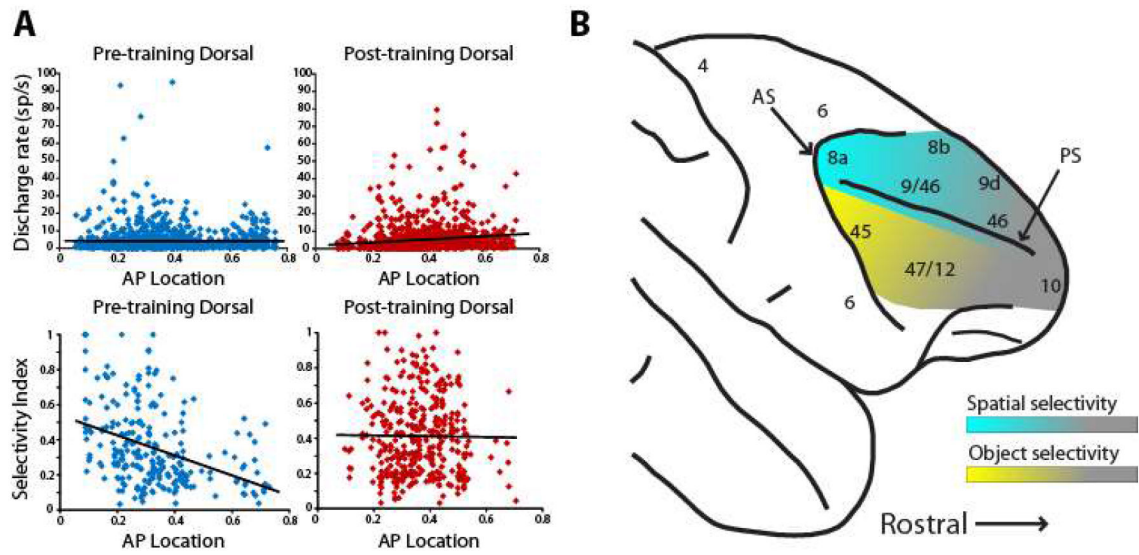


Figure 6. Stimulus selectivity along the dorso-ventral and anterior-posterior axes of the PFC before and after training. **A.** Firing rate and selectivity of dorsal prefrontal neurons at different positions along the anterior-posterior axis to spatial stimuli, before and after training. **B.** Schematic illustration of spatial and object selectivity across the PFC. Panel A from Riley et al., 2018; panel B from Constantinidis and Qi al. 2018.

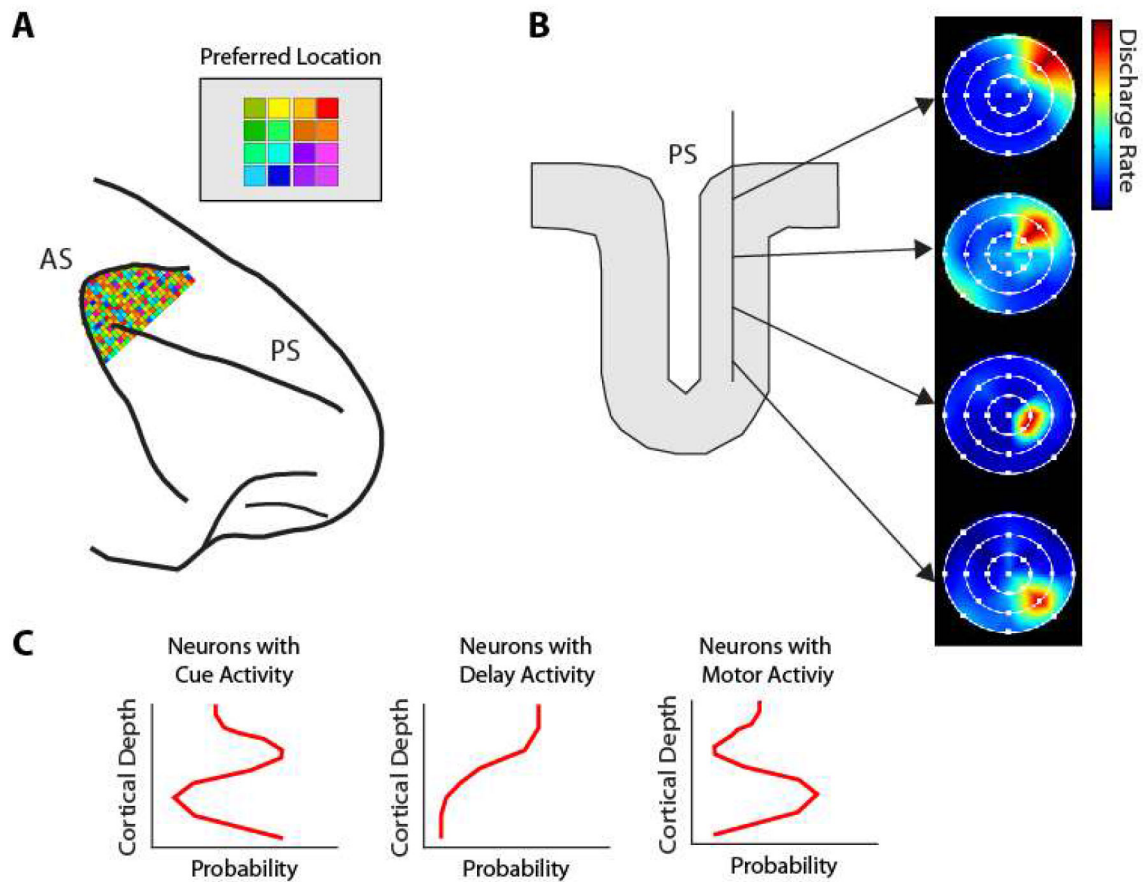


Figure 7. Local organization of the prefrontal cortex. **A.** No retinotopic map of visual space is present in dorsal posterior PFC, whose neurons do display strong selectivity for space. **B.** Some local organization and orderly progression of receptive fields is revealed in tangential electrode penetrations in the principal sulcus. **C.** Organization of activity across the depth of the PFC. Cue responses predominate in middle layers; delay period activity in upper layers; response activity in deeper layers of the PFC. Panel A based on Leavitt et al., 2018; panel B based on Arnsten et al., 2013; panel C based on Markowitz et al., 2015.

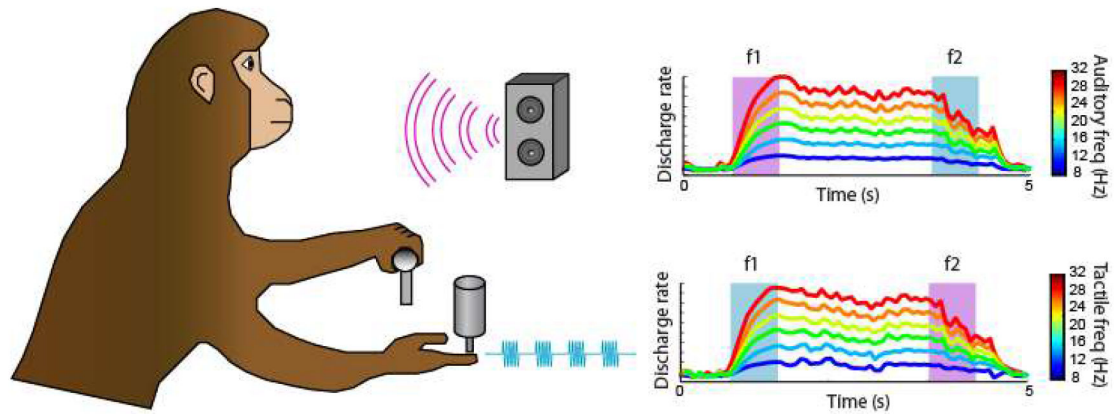


Figure 8.

Schematic illustration of a cross-modal WM task. The monkey needs to remember the frequency of either a tactile or auditory stimulus and compare it with a stimulus of either modality. Right panels represent schematically the responses of a single neuron in the pre-Supplementary Motor Area for an auditory stimulus (purple bar) followed by a tactile stimulus (blue bar), or vice versa (bottom). Responses of the neuron during the delay period were modulated monotonically depending on the frequency of the stimulus, adhering to the same monotonic code of firing rate as a function of frequency for both the remembered auditory and tactile stimuli. From Constantinidis, 2016, referencing the data of Vergara et al., 2016.

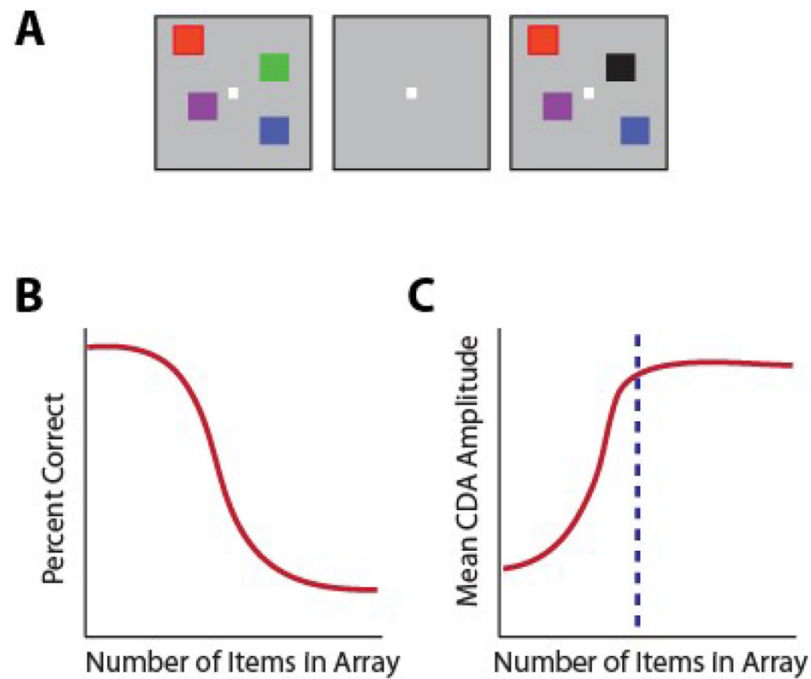


Figure 9.

Properties of working memory capacity in human subjects. **A.** Change detection paradigm. Subjects need to maintain in memory a set of colored squares and report if a second display was identical to the first or not. **B.** Performance on the task declines as a function of set size. **C.** Mean Contralateral delay activity measured from posterior EEG electrodes as a function of elements in the array. Peak CDA amplitude predicts capacity (vertical) line. Based on Fukuda et al., 2010.

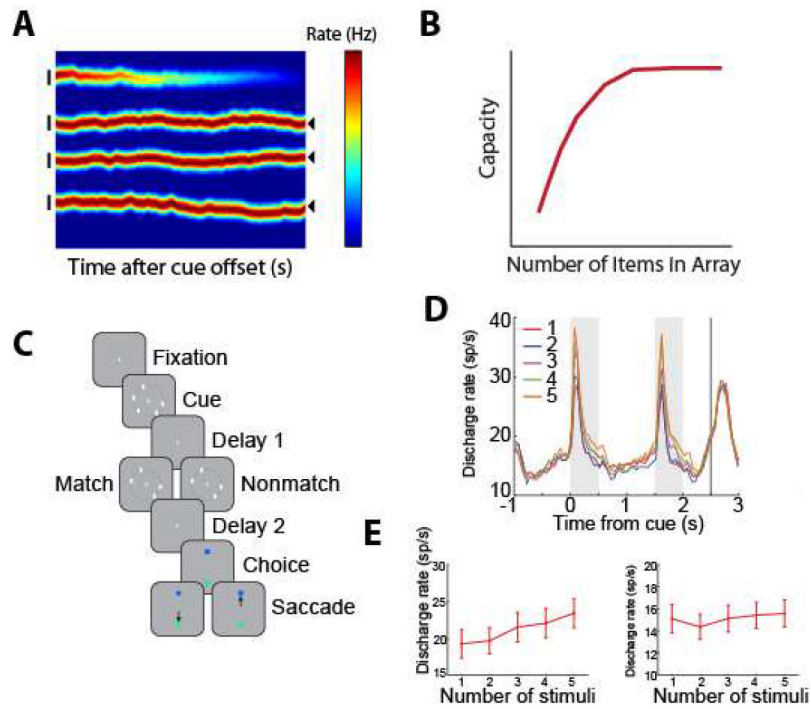


Figure 10.
A. Predicted model of working memory capacity in the context of the bump attractor. Multiple stimuli are encoded at the beginning of the trial. An item that decays, fails to be recalled at the end of the trial. **B.** Capacity curves of monkeys. **C.** WM capacity task. **D.** Population firing rate for displays with different numbers of stimuli. **E.** Average firing rate during the cue and delay period. Adapted from Tang et al., 2019.

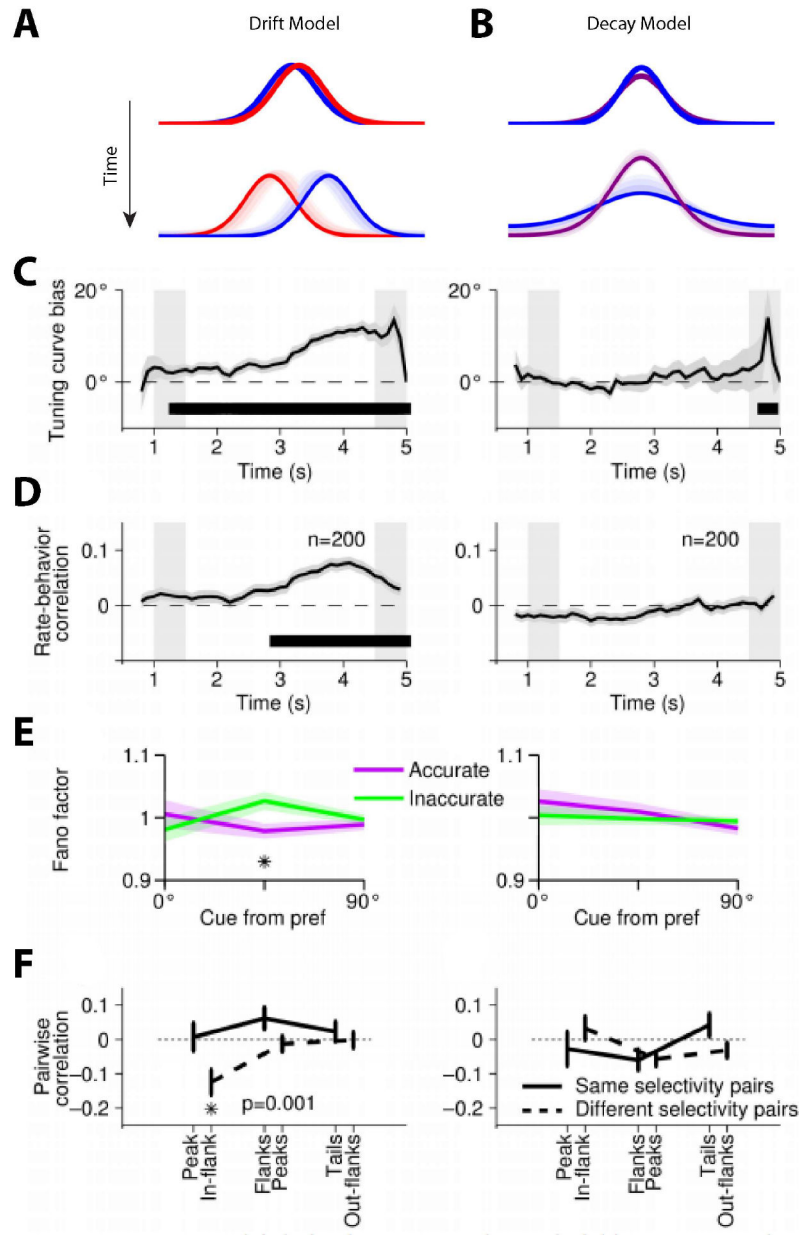


Figure 11. Predictions of **A**, the bump attractor model suggesting that drift is the main source of imprecision in working memory vs. **B**, a model where the bump of activity decays. **C**. For the bump attractor, but not the decaying bump model, the tuning curve bias computed from trials with clockwise versus counterclockwise deviations becomes increasingly positive through the delay. **D**. Only for the bump attractor model, spike counts in response to flank stimuli correlated with behavioral deviations in the direction of the neuron’s preferred cue as delay progressed. **E**. For the bump attractor model, but not the decaying bump model, Fano Factors computed separately for trials when a flank stimulus was presented are larger when considering inaccurate trials with large behavioral deviations (solid line) as compared to accurate trials (dashed line). **F**. Only for the bump attractor model, neuron pair noise

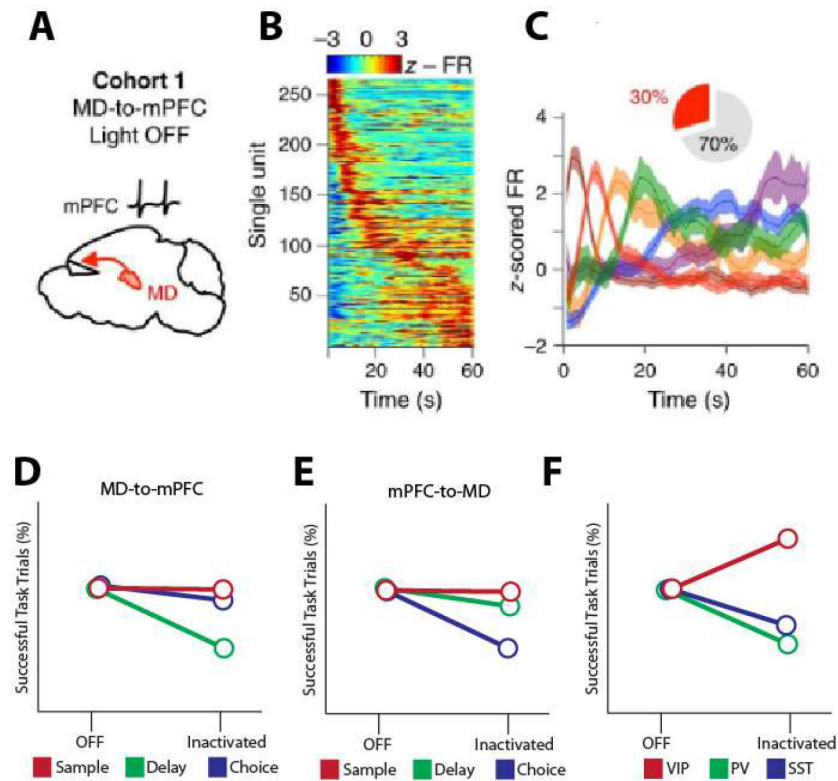
correlation is negative for those pairs with dissimilar tuning, when responding to a middle flank stimulus. From Wimmer et al., 2014.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Figure 12.**

A. Patterns of activation of mouse working memory, recorded from mPFC. **B.** Individual neurons are typically active over a short period. **C.** Mean z-scored firing rate of delay-elevated units identified after clustering into six groups based on temporal correlation in firing rates from dark red (earliest) to purple (latest). Approximately 30% of all mPFC neurons in the dataset exhibited significant delay-elevated activity (inset). **D.** Behavioral effects of inactivation of MD-to-mPFC terminals. **E.** Behavioral effects of inactivation of mPFC-to-MD terminals. **F.** Behavioral effects of inactivation of different mPFC interneuron populations during an auditory working memory task. Panels A-E from Bolkan et al., 2017. Panel F is based on Kamigaki and Dan, 2017.

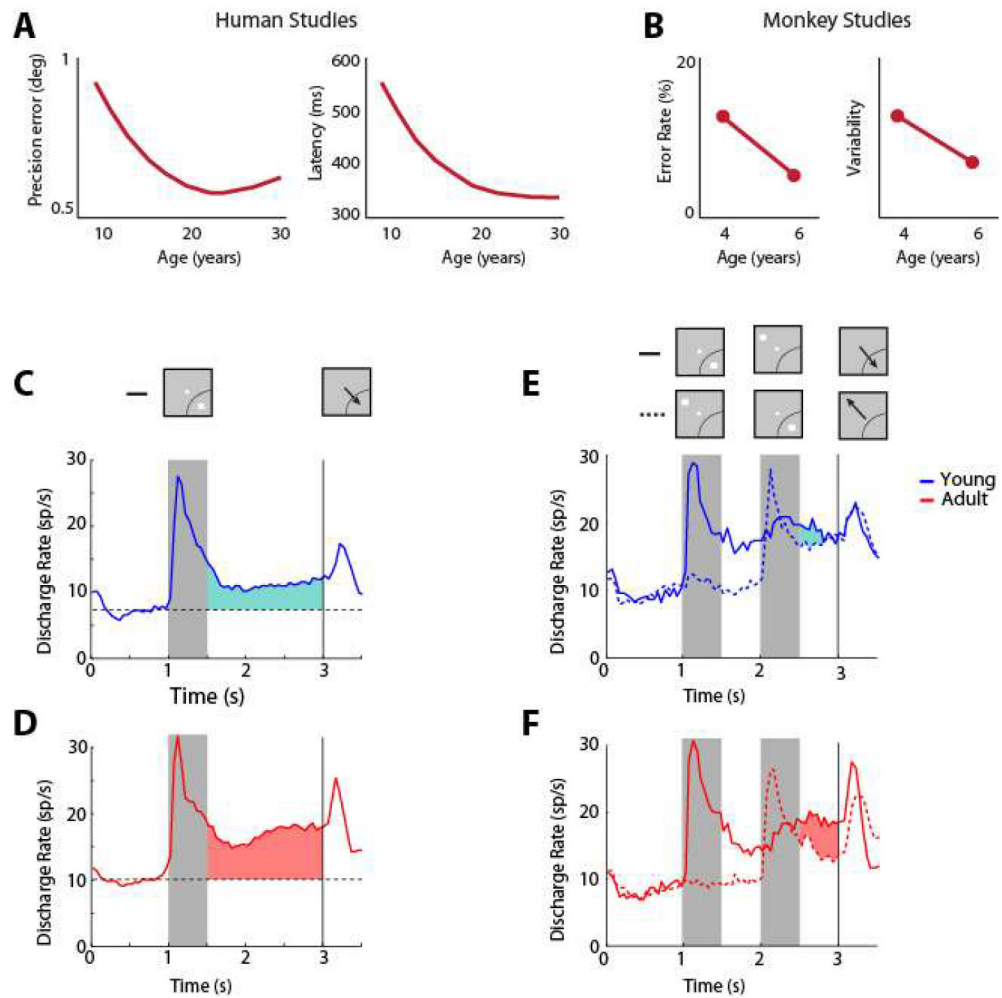


Figure 13. **A.** Behavioral improvement of human working memory in adolescence. **B.** Behavioral improvement of monkey WM in adolescence. **C.** Persistent discharges in the adolescent and **D,** adult PFC in the Oculomotor Delayed Response task. **E-F.** Activity in the adolescent and adult PFC in the distractor task. Panel A based on Simmons et al., 2017; panels B-F based on Zhou et al., 2016.

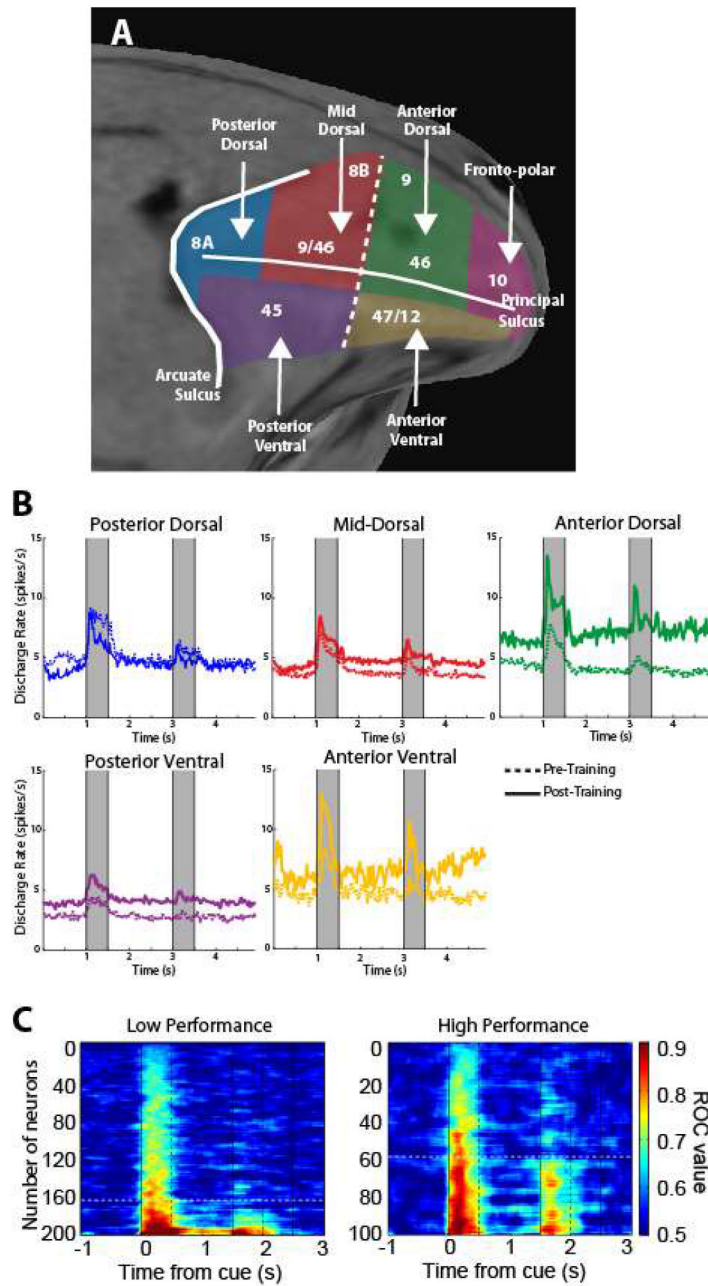


Figure 14.

A. Anatomical MRI of the monkey lateral prefrontal cortex with anterior/posterior and dorsal/ventral subdivisions indicated relative to the Principle and Arcuate Sulci. **B.** Mean firing rate of neurons recorded in these subdivisions in monkeys both before and after they were trained to perform spatial WM tasks. Gray bars represent stimulus presentations. Data are shown separately for each prefrontal region. **C.** ROC analysis for recordings from low-performance and high-performance sessions, as monkeys were trained in a multi-stimulus WM task. Panels A and B, based on Riley et al., 2018. Panel C from Tang et al., 2019.